UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

Testing for Time Stochastic Dominance

Kyungho Lee	Oliver Linton	Yoon-Jae Whang
Seoul National University	University of Cambridge	Seoul National University

Abstract

We propose nonparametric tests for the null hypothesis of time stochastic dominance. Time stochastic dominance makes a partial order of different prospects over time based on the net present value criteria for general utility and time discount function classes. For example, time stochastic dominance can be used for ranking investment strategies or environmental policies based on the expected net present value of the future benefits. We consider an Lp integrated test statistic and derive its large sample distribution. We suggest a path-wise bootstrap procedures that allows for time dependence in a panel data structure. In addition to the least favourable case based bootstrap method, we describe two approaches, the contact-set approach and the numerical delta method, for the purpose of enhancing a power of the test. We prove the asymptotic validity of our testing procedures. We investigate the finite sample performance of the tests in simulation studies. As an illustration, we apply the proposed tests to evaluate the welfare improvement of the Thailand's Million Baht Village Fund Program.

Reference Details

20121	Cambridge Working Papers in Economics
2020/53	Cambridge-INET Working Paper Series
Published	10 December 2020
Key Words	C10, C12, C14
JEL Codes	Bootstrap, Discounting, Stochastic Dominance, Testing
Websites	www.econ.cam.ac.uk/cwpe www.inet.econ.cam.ac.uk/working-papers

TESTING FOR TIME STOCHASTIC DOMINANCE

KYUNGHO LEE, OLIVER LINTON, AND YOON-JAE WHANG

Abstract. We propose nonparametric tests for the null hypothesis of time stochastic dominance. Time stochastic dominance makes a partial order of different prospects over time based on the net present value criteria for general utility and time discount function classes. For example, time stochastic dominance can be used for ranking investment strategies or environmental policies based on the expected net present value of the future benefits. We consider an L_p integrated test statistic and derive its large sample distribution. We suggest a path-wise bootstrap procedures that allows for time dependence in a panel data structure. In addition to the least favorable case based bootstrap method, we describe two approaches, the contact-set approach and the numerical delta method, for the purpose of enhancing a power of the test. We prove the asymptotic validity of our testing procedures. We investigate the finite sample performance of the tests in simulation studies. As an illustration, we apply the proposed tests to evaluate the welfare improvement of the Thailand's Million Baht Village Fund Program.

1. Introduction

Decisionmakers and academics often have to compare different projects that contain both risk (uncertain outcomes) and outcomes realized at different time periods. Examples include investment decisions, environmental policies, a microfinance program for improvement of living standards, an R&D investment aimed at increasing in productivity, a human-capital investment and numerous others. In the intertemporal choice literature, Net Present Value (NPV) is the cornerstone of comparing dynamic welfare outcomes and for making an ordering of the prospects. Most analyses depend on specific parametric utility functions to account for risk and specific time-discounting when comparing NPV's. This is problematic for three reasons. First, comparisons are usually sensitive to assumptions about risk preferences, Mehra and Prescott (1985). Second, specifying the discounting function is also important but controversial, see Cohen, Ericson, Laibson,

Date: December 9th, 2020.
Whang's research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5A2A1044600). Lee acknowledges funding from the BK21Plus Program of the Ministry of Education and National Research Foundation of Korea (NRF-21B20130000013).

and White (2020), and can lead to different orderings.[1] Finally, specific utility and time-discounting assumptions only reflect a very small part of potential preferences, so forging a general consensus on an ordering, and hence the appropriate course of action, cannot be achieved.

Dietz and Matei (2016) introduced the Time Stochastic Dominance (TSD) concept as a general framework for comparing choices with both timing elements and risk. This approach combines the pure time specific Stochastic Dominance (SD) framework, Levy (1973, 1998), with the Time Dominance paradigm, Ekern (1981). TSD provides a partial order to rank different distributions over time by comparing their NPV under general assumptions on both the utility function and the time discount function. By allowing a general class of utility functions, which may contain the widely used CRRA class, and a general class of time discount functions, which may contain the widely used discount functions such as the exponential or the hyperbolic discount functions, TSD allows for a consensus view among economic agents with a general preference structure over uncertain portfolios or policies.

The goal of this paper is to develop a nonparametric testing procedure for the null hypothesis of time stochastic dominance. Despite its usefulness, to the best of our knowledge, a formal statistical inference method for TSD has not been available in the literature. This paper considers an L_p-type test statistic and suggests bootstrap methods to compute critical values and proves their asymptotic validity.

Our paper contributes to a big literature on SD testing. McFadden (1989) introduces Kolmogorov-Smirnov type nonparametric test of first and second order SD hypotheses. After the pioneering work of McFadden (1989), Klecan, McFadden and McFadden (1991), Kaur, Prakasa Rao, and Singh (1994), Anderson (1996) and Davidson and Duclos (2000) also propose different approaches to SD testing. Barrett and Donald (2003) suggest a consistent bootstrap method to test an arbitrary SD order between two prospects under the assumption of i.i.d data and mutually independent prospects or outcomes. Linton, Maasoumi, and Whang (2005, hereafter LMW) show a consistent subsampling method testing for SD under general sampling scheme, which allows time series dependency in the data and mutual dependence between the outcomes as well as allowing for estimated finite-dimensional parameters. Linton, Song and Whang (2010, hereafter LSW) propose

[1]For example, Stern (2007), Nordhaus (2007) and Weitzman (2007) debate sensitivity of discounting factor assumptions in a context of assessing environmental damages.

an improved inference method for SD tests based on contact set estimation, at the same time they allowed for nonparametric and parametric components in the DGP. Whang (2019) contains a comprehensive survey of the SD testing literature. However, none of these papers addresses the issue raised by time discounting, which means the economic interpretation of any ordering is confined to static comparisons.

Our paper describes three methods for calculating a critical value. We first suggest the conventional approach for estimating critical values that imposes the least favorable case (LFC) of the null hypothesis to mimic the asymptotic distribution of the test statistic. However, imposing LFC for every time period is too conservative in our dynamic setting. To overcome it, we estimate the contact set, where the asymptotic distribution of the test statistic does not degenerate under the null hypothesis, to imitate the limiting distribution more directly. In addition, we describe the numerical delta method suggested by Hong and Li (2018) as an alternative procedure.

Our sampling scheme is suitable for a panel data structure. We employ a bootstrap procedure that resamples a time path in order to allow for *general individual time dependence.* Because panel data is essential for catching individual dynamics, our test is widely applicable for various empirical analyses. For instance, in the development economics literature, many researchers are interested in a certain policy's long-run impact, such as the welfare improvement, so they employ data that tracks individuals or households over a long period.

Our paper considers the case where prospects depend on unknown parameters. For example, residuals from some estimated model can be a testing object. Testing for *residual TSD* is very useful for policy makers to concentrate on an outcome of interest because it enables them to control systematic differences between prospects. For instance, the effect of a subsidy policy for electric vehicles to lower regional air pollution level depends on certain local characteristics, such as power plant location, source of electricity generation, and an automobile composition in town.[2] In this case, the residuals from regressing assessed regional air pollution damage on such covariates are necessary for comparing the environmental benefits of two policies.

We conduct Monte Carlo simulations to evaluate the finite sample performance of our proposed methods. In addition to the power and size-control property, our particular

[2]Holland et al. (2016) discuss the importance of local factors on the environmental policy evaluation.

interest is the impact of the time order on the testing results. As the time order becomes higher, the set of compatible time preferences become narrower i.e., the time preference requires more conditions. We evaluate whether restricting preferences are well reflected in our testing.

As an illustration, we apply our testing to evaluate the impact of Thailand's Million Baht Village Fund Program on dynamic welfare. By using our test, policy makers can have a general consensus on a policy that pursues a long-term goal such as development of the local economy.

The rest of this paper is organized as follows: In Section 2, we give definitions of TSD, and introduce hypotheses of interest. In Section 3, we suggest test statistics and prove its asymptotic properties. In Section 4, we give bootstrap inference procedures and prove asymptotic validity. In Section 5, we describe numerical delta method suggested by Hong and Li (2018) as an alternative inference method. In Section 6, we conduct Monte Carlo experiments for evaluating finite sample performance. In Section 7, as an illustration, we evaluate the welfare improvement of Thailand's Million Baht Village Fund Program. Concluding remarks are in Section 8.

2. Time Stochastic Dominance and The Hypotheses of Interest

2.1. **Time Stochastic Dominance.** Let $X_1 := \{X_{1t} : t \in \mathcal{T}\}$ and $X_2 := \{X_{2t} : t \in \mathcal{T}\}$ be two prospects that yield random cash flows realizations over time. We regard them as stochastic processes indexed by $t \in \mathcal{T}$. For simplicity, we assume that X_{1t} and X_{2t} have the common support $\mathcal{X} = [\underline{x}, \bar{x}]$ for all $t \in \mathcal{T}$.[3] The concept of TSD can be defined both in continuous time with $\mathcal{T} = [0, T]$, and in discrete time with $\mathcal{T} = \{0, 1, .., T\}$, see Ekern (1981) and Dietz and Matei (2016) for details. For brevity, we introduce the concept in the case of discrete time.

Let $f_k(\cdot, t)$ and $F_k(\cdot, t) := \int_{\underline{x}}^{\cdot} f_k(z, t)dz$ denote the density and distribution function, respectively, of X_{kt} for $k = 1, 2$ and $t \in \mathcal{T}$. Let $u : \mathcal{X} \mapsto \mathbb{R}$ be a utility function and $v : \mathcal{T} \mapsto \mathbb{R}$ be a time-discount function. Both functions v and u are assumed to be continuously differentiable. The expected discounted utility of the prospect X_k at $t = 0$

[3]The assumption of compact support is for simplicity. We may allow it to be unbounded by introducing a weight function in the definition of the test statistics discussed below, see Linton, Song and Whang (2010). Also, the support may be allowed to depend on t.

is given by

$$NPV_{v,u}(X_k) := \sum_{t=0}^{T} v(s)\mathbf{E}_F u(X_{kt}) = \sum_{t=0}^{T} v(t)\left[\int_{\mathcal{X}} u(x) f_k(x,t)dx\right],$$

which depends on the flow utility u, the discount function v and the distribution of the outcomes. Individuals are assumed to rank projects or outcomes rationally according to the values of $NPV_{v,u}(X_k)$, $k = 1, 2$.

Define the nested classes of utility functions $\mathcal{U}_1 = \{u : u^{(1)}(x) \geq 0\}$ and $\mathcal{U}_2 = \{u : u \in \mathcal{U}_1, u^{(2)}(x) \leq 0\}$, where $u^{(s)}$, $s \in \mathbb{Z}^+$ denote the s-th order derivative of u. More generally, the higher-order utility function classes are defined recursively as $\mathcal{U}_m = \{u : u \in \mathcal{U}_{m-1}, (-1)^m u^{(m)}(x) \leq 0\}$ for $m \geq 2$. Note that $\mathcal{U}_1$ corresponds to the classes of monotonically increasing utility functions, while $\mathcal{U}_2$ corresponds to those of monotonically increasing and concave utility functions associated with risk averse behavior. The commonly used CRRA class is in $\mathcal{U}_2$.

Next, we define classes of discount functions. Let $v^0(t) = v(t)$ and $v^n(t) = v^{n-1}(t + 1) - v^{n-1}(t)$ for $n \in \mathbb{Z}^+$. Define $\mathcal{V}_0 = \{v : v(0) = 1, v(t) \geq 0 \text{ for } t = 0, \ldots, T\}$, $\mathcal{V}_1 = \{v : v \in \mathcal{V}_0, v(t+1) - v(t) \leq 0 \text{ for } t = 0, \ldots, T-1\}$, and $\mathcal{V}_2 = \{v : v \in \mathcal{V}_1, v(t+2) - v(t+1) \geq v(t+1) - v(t) \text{ for } t = 0, \ldots, T-2\}$. More generally, define $\mathcal{V}_n = \{v : v \in \mathcal{V}_{n-1}, (-1)^n v^n(t) > 0\}$ for $n \geq 1$. The class $\mathcal{V}_0$ consists of all strictly positive discount functions, which is not restrictive because there is always some positive degree of time preference, however small. The class $\mathcal{V}_1$ comprises strictly decreasing discount functions, representing impatience over time. On the other hand, $\mathcal{V}_2$ is the class of strictly decreasing and convex discount functions, according to which impatience decreases over time. Note that $\mathcal{V}_2$ contains exponential and hyperbolic discount functions, both of which are widely used throughout the economics literature.

Define the differences

$$D^{(1,1)}(x,t) := F_1^{(1,1)}(x,t) - F_2^{(1,1)}(x,t), \text{ where}$$
$$F_k^{(1,1)}(x,t) := \textstyle\sum_{s=0}^{t} F_k(x,s) = \sum_{s=0}^{t} \int_{\underline{x}}^{x} f_k(z,s)dz,$$

for $k = 1, 2$ and $(x,t) \in \mathcal{X} \times \mathcal{T}$. The next definition relates the unobservable utility based comparison to a comparison of the observable distributions.

Definition 1. X_1 *First order Time and First order Stochastic Dominates* X_2, denoted as $X_1 \succeq_{1T1SD} X_2$, if and only if,

(a) $NPV_{v,u}(X_1) - NPV_{v,u}(X_2) \geq 0, \forall (v,u) \in \mathcal{V}_1 \times \mathcal{U}_1$, or

(b) $D^{(1,1)}(x,t) \leq 0$, $\forall (x,t) \in \mathcal{X} \times \mathcal{T}$.

To denote higher order of time stochastic dominance, for integers $m \geq 2, n \geq 1$ and $k = 1, 2$, define the following integral:

$$F_k^{(n,m)}(x,t) = \int_{\underline{x}}^{x} F_k^{(n,m-1)}(z,t)dz = \sum_{s=0}^{t} F_k^{(n-1,m)}(x,s) = \sum_{s=0}^{t} \int_{\underline{x}}^{x} F_k^{(n-1,m-1)}(z,s)dz$$

for $(x,t) \in \mathcal{X} \times \mathcal{T}$, with the convention that $F_k^{(0,1)} = F_k$. Using integration by parts, we can establish the following result:

$$F_k^{(n,m)}(x,t) := \sum_{s=0}^{t} (t-s+1)^{n-1} \int_{\underline{x}}^{x} \frac{(x-z)^{m-1} 1\{z \leq x\}}{(m-1)!} dF_k(z,s). \tag{2.1}$$

Define

$$D^{(n,m)}(x,t) := F_1^{(n,m)}(x,t) - F_2^{(n,m)}(x,t)$$

for $(x,t) \in \mathcal{X} \times \mathcal{T}$. Definition 1 can be generalized to higher-order dominance (Dietz and Matei (2016, Proposition 5)):

Definition 2. X_1 *n-th order Time and m-th order Stochastic Dominates* X_2, denoted as $X_1 \succeq_{nTmSD} X_2$, if and only if,

(a) $NPV_{v,u}(X_1) \geq NPV_{v,u}(X_2)$, $\forall (v,u) \in \mathcal{V}_n \times \mathcal{U}_m$, or

(b) (i) $D^{(i+1,j+1)}(\bar{x}, T) \leq 0$; (ii) $D^{(n,j+1)}(\bar{x}, t) \leq 0$ $\forall t \in \mathcal{T}$; (iii) $D^{(i+1,m)}(x,T) \leq 0$ $\forall x \in \mathcal{X}$; (iv) $D^{(n,m)}(x,t) \leq 0$ $\forall (x,t) \in \mathcal{X} \times \mathcal{T}$, where $i \in \{0, \ldots, n-1\}$ and $j \in \{1, \ldots, m-1\}$.

For example, when $n = 1$ and $m = 2$, then Definition 2(b) amounts to $D^{(1,2)}(x,t) \leq 0$ for all $(x,t) \in \mathcal{X} \times \mathcal{T}$. Also, when $n = m = 2$, it corresponds to $D^{(1,2)}(x,T) \leq 0$ for all $x \in \mathcal{X}$ and $D^{(2,2)}(x,t) \leq 0$ for all $(x,t) \in \mathcal{X} \times \mathcal{T}$. Note that the latter case has an additional requirement: $D^{(1,2)}(x,T) \leq 0$ for all $x \in \mathcal{X}$ at the terminal period T. This definition is the basis of our testing strategy. Specifically, we replace 2(a) the hypothesis of interest by 2(b) the equivalent formulation.

2.2. **The Hypotheses of Interest.** The null hypothesis of n-order time and m-th order stochastic dominance is given by

$$H_0^{(n,m)} : NPV_{v,u}(X_1) \geq NPV_{v,u}(X_2),\ \forall (v,u) \in \mathcal{V}_n \times \mathcal{U}_m, \tag{2.2}$$

which is equivalent to F_1 and F_2 satisfying Definition 2(b). For example, $H_0^{(2,2)}$ says that all risk averse individuals who discount with decreasing and convex discount functions would prefer project one to project two. The alternative hypothesis $H_1^{(n,m)}$ is the negation of $H_0^{(n,m)}$, that is, there exists at least one person with $v, u \in \mathcal{V}_n \times \mathcal{U}_m$ who ranks the prospects differently.

Under the discrete time setup $\mathcal{T} = \{0, 1, \ldots, T\}$, Definition 2(b) consists of $L := (n + T) \times (m-1)$ inequalities for the functional $D^{(i+1,j+1)}(\cdot, t) := F_1^{(i+1,j+1)}(\cdot, t) - F_2^{(i+1,j+1)}(\cdot, t)$ for $i \in \{0, \ldots, n-1\}, j \in \{1, \ldots, m-1\}, t \in \{0, \ldots, T\}$. Below, for notational convenience, we shall denote the functionals in Definition 2(b) as $v_l(\cdot)$, $l = 1, \ldots, L$, i.e.,

$$v_l(x) := \begin{cases} D^{(1,l+1)}(x,T), & 1 \leq l \leq (m-1) \\ D^{(2,l+1-(m-1))}(x,T), & (m-1)+1 \leq l \leq 2(m-1) \\ \vdots & \vdots \\ D^{(n,l+1-(n-1)(m-1))}(x,T), & (n-1)(m-1)+1 \leq l \leq n(m-1) \\ D^{(n,l+1-n(m-1))}(x,0), & n(m-1)+1 \leq l \leq (n+1)(m-1) \\ D^{(n,l+1-(n+1)(m-1))}(x,1), & (n+1)(m-1)+1 \leq l \leq (n+2)(m-1) \\ \vdots & \vdots \\ D^{(n,l+1-(n+T-1)(m-1))}(x,T-1), & (n+T-1)(m-1)+1 \leq l \leq (n+T)(m-1). \end{cases} \tag{2.3}$$

Let $\Lambda_p : \mathbb{R}^L \to [0, \infty)$ be a nonnegative and increasing function for $p \in \{1, 2\}$. Specifically, we focus on the following map:

$$\Lambda_p(v_1, \ldots, v_L) = \sum_{l=1}^{L} [v_l]_+^p, \text{ or } \left(\max\{[v_1]_+, \ldots, [v_L]_+\}\right)^p, \tag{2.4}$$

where for $a \in \mathbb{R}$, $[a]_+ = \max\{a, 0\}$. Define the population quantity

$$d^* = \int_{\mathcal{X}} \Lambda_p\left(v_1(x), \ldots, v_L(x)\right) dx. \tag{2.5}$$

Then, the hypotheses of interest can be equivalently stated as

$$H_0^{(n,m)} : d^* = 0 \text{ vs. } H_1^{(n,m)} : d^* > 0.$$

The test statistic introduced below is based on the sample analogue of d^*.

We allow for the case where X_{kt} depends on unknown parameters, so that we may control for systematic differences using a model specification. Specifically, we let $X_{kt}(\theta)$ be specified as

$$X_{kt}(\theta) = \varphi_{kt}(W_t, \theta),\ k = 1, 2,$$

where W_t is a random vector in $\mathbb{R}^{d_W}$ and $\varphi_{kt}(\cdot, \theta)$ is a real-valued function known up to the parameter $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$. We let $X_{kt} = X_{kt}(\theta_0)$ for some $\theta_0 \in \Theta$. For example, we may take X_{kt} to be the residual from the regression

$$X_{kt} = Y_{kt} - Z_{kt}^{\intercal}\theta_0,$$

where $Y_{kt} = Z_{kt}^{\intercal}\theta_0 + \epsilon_{kt}$ with $\mathbf{E}(\epsilon_{kt}|Z_{kt}) = 0$ a.s. In this case, we take $W = (Y, Z)$ and $\varphi_{kt}(w, \theta) = y_{kt} - z_{kt}^{\intercal}\theta$, $w = (y, z)$.

3. Test Statistics

We now define our test statistic based on data $\{W_{kti},\ i = 1, \ldots, N_k,\ t \in \mathcal{T},\ k = 1, 2\}$. For $k = 1, 2$, let

$$\bar{F}_k(x, t, \theta) := \frac{1}{N_k}\sum_{i=1}^{N_k} 1(X_{kti}(\theta) \le x) \text{ and } \bar{F}_k^{(1,1)}(x, t, \theta) := \sum_{s=0}^{t} \bar{F}_k(x, s, \theta)$$

denote the empirical distribution function (EDF) and the EDF with time-accumulation (EDFT), respectively. Likewise, define the empirical analogue of the general integrated CDF with time-accumulation as

$$\begin{aligned}\bar{F}_k^{(n,m)}(x, t, \theta) =& \frac{1}{N_k}\sum_{i=1}^{N_k}\sum_{s=0}^{t} \frac{(t - s + 1)^{n-1}(x - X_{ksi}(\theta))^{m-1}1(X_{ksi}(\theta) \le x)}{(m-1)!} \\ =& \frac{1}{N_k}\sum_{i=1}^{N_k}\sum_{s=0}^{t} a_t(s)h_x(X_{ksi}(\theta)),\end{aligned} \tag{3.1}$$

where $a_t(s) := (t - s + 1)^{n-1}$ and for $m \ge 1$, $h_x(\varphi) := (x - \varphi)^{m-1}1\{\varphi \le x\}/(m-1)!$ Let $\bar{D}^{(n,m)}(x, t, \hat{\theta}) := \bar{F}_1^{(n,m)}(x, t, \hat{\theta}) - \bar{F}_2^{(n,m)}(x, t, \hat{\theta})$, where $\hat{\theta}$ denotes a consistent estimator of θ_0. Let $\hat{v}_l(x)$ be the sample analogue of $v_l(x)$ (defined in (2.3)), $l = 1, \ldots, L$, with $D^{(i,j)}(x, t)$ replaced by $\bar{D}^{(i,j)}(x, t, \hat{\theta})$ for $i \in \{1, \ldots, n\}$, $j \in \{2, \ldots, m\}$ and $t \in \{0, \ldots, T\}$.

To test the null hypothesis $H_0^{(n,m)}$, we consider the following one-sided L_p-type test statistic (based on the sample analogue of d^* defined in (2.5)):

$$T_N = r_N^p \int_{\mathcal{X}} \Lambda_p \left(\hat{v}_1(x), \ldots, \hat{v}_L(x)\right) dx, \tag{3.2}$$

where $r_N := \sqrt{(N_1 N_2)/(N_1 + N_2)}$.

We next present the regularity conditions that we work with to derive the properties of T_N. In the following assumptions, we specify the conditions for the data generating process of W and the map φ_{kt}. Let $\mathcal{P}$ be the collection of all the potential distributions of W that satisfy Assumptions 1, 2, and 3 below. Let $B_\Theta(\delta) \equiv \{\theta \in \Theta : ||\theta - \theta_0|| < \delta\}$ be the δ-neighborhood of θ_0, where $|| \cdot ||$ denotes the Euclidean norm.

Assumption 1.

(a) $\{W_{kti}\}_{i=1}^{N_k}$ is a random sample for $t \in \mathcal{T}$ and $k = 1, 2$.

(b) As $N_1, N_2 \to \infty$, $N_1/(N_1 + N_2) \to \lambda \in (0, 1)$ for $t \in \mathcal{T}$.

(c) For some $\delta > 0$, $\sup_{P \in \mathcal{P}} \mathbf{E}_P[\sup_{\theta \in B_\Theta(\delta)} |X_{kti}(\theta)|^{2((m-1)\vee 1)+\delta}] < \infty$ for $t \in \mathcal{T}$ and $k = 1, 2$.

(d) For some $\delta > 0$, there exists a non-random $d_\theta \times 1$ vector $\Gamma_{kt,P}(x)$ such that

$$\begin{aligned} & |\mathbf{E}_P \left[h_x(X_{kti}(\theta))\right] - \mathbf{E}_P \left[h_x(X_{kti}(\theta_0))\right] - \Gamma_{kt,P}(x)^\intercal (\theta - \theta_0)| \\ \leq \;& C||\theta - \theta_0||^2, \; t \in \mathcal{T}, \; k = 1, 2, \end{aligned}$$

with constant C does not depend on P.

Assumption 1(a) implies that, for each $k = 1, 2$, the observations are independent across i, but are possibly dependent over time t for a given i. This assumption allows X_{1ti} and X_{2ti} to be dependent for any (t, i). Assumption 1(b) is a moment condition with local uniform boundedness. In the example of linear regression models where $Y_{kti} = Z_{kti}^\intercal \theta_0 + \epsilon_{kti}$, we may write $X_{kti}(\theta) = \epsilon_{kti} + Z_{kti}^\intercal(\theta_0 - \theta)$ and hence the condition is satisfied when $\sup_{P \in \mathcal{P}} \mathbf{E}_P[|\epsilon_{kti}|^{2((m-1)\vee 1)+\delta}] < \infty$ and $\sup_{P \in \mathcal{P}} \mathbf{E}_P[|Z_{kti}|^{2((m-1)\vee 1)+\delta}] < \infty$. Assumption 1(c) is differentiability of the functional $\int h_x(X_{kti}(\theta))dP$ in $\theta \in B_\Theta(\delta)$. When $m = 1$, in the example of linear regression models, it is satisfied with $\Gamma_{kt,P}(x) = \mathbf{E}_P f_{\epsilon|Z}(x|Z_{kti})Z_{kti}$ when the conditional density $f_{\epsilon|Z}(\cdot|Z_{kti})$ of ϵ_{kti} given Z_{kti} is second order continuously differentiable with bounded derivative and the moment condition in 2(b) holds. When

$m = 2$, $h_x(\varphi)$ is Lipschitz in φ with the coefficient bounded by $C|x-\varphi|^{m-2}$. Hence, Assumption 1(c) follows if the moment condition in Assumption 1(b) is satisfied.

Assumption 2.

(a) $X_{kti}(\theta_0)$ has distribution function $F_k(\cdot,t)$ and has density $f_k(\cdot,t)$ with respect to Lebesgue measure for $t \in \mathcal{T}$ and $k = 1, 2$.

(b) Condition (A) below holds when $m = 1$ and condition (B) holds when $m = 2$:

(A) There exist $\delta, C > 0$ and a subvector W_1 of W such that (i) the conditional density of W given W_1 is bounded uniformly over $\theta \in B_\Theta(\delta)$ and over $P \in \mathcal{P}$, (ii) for each θ_1 and θ_2 in $B_\Theta(\delta)$, $\varphi_{kt}(W,\theta_1) - \varphi_{kt}(W;\theta_2)$ is measurable with respect to the σ-field of W_1, and (iii) for each $\theta \in B_\Theta(\delta)$ and for each $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}} \sup_{w_1} \mathbf{E}_P \left[\sup_{\theta_1 \in B_\Theta(\delta)} |\varphi_{kt}(W,\theta_1) - \varphi_{kt}(W,\theta)|^2 \, |W_1 = w_1 \right] \le C\varepsilon^{2s_2} \tag{3.3}$$

for some $s_2 \in (\lambda/2, 1]$ with $\lambda = 2 \times 1\{m=1\} + 1\{m>1\}$, where the supremum over w_1 runs in the support of W_1.

(B) There exist $\delta, C > 0$ such that Condition (iii) above is satisfied with the conditional expectation replaced by the unconditional one.

We assume that $\hat{\theta}$ satisfies the following conditions:

Assumption 3.

(a) For each $\varepsilon > 0$, $\sup_{P\in\mathcal{P}} P\{||\hat{\theta}-\theta_0|| > \varepsilon\} = o(1)$ as $N_1, N_2 \to \infty$.

(b) For each $\varepsilon > 0, k = 1, 2$ and $t \in \mathcal{T}$,

$$\sup_{P\in\mathcal{P}} P \left\{ \sup_{x\in\mathcal{X}} \left| \sqrt{N_k}\Gamma_{kt,P}(x)'[\hat{\theta}-\theta_0] - \frac{1}{\sqrt{N_k}} \sum_{i=1}^{N_k} \psi_{x,kt}(W_i;\theta_0) \right| > \varepsilon \right\} \to 0, \tag{3.4}$$

where $\psi_{x,kt}(\cdot)$ satisfies that there exist $\eta, \delta > 0$ such that for all $x \in \mathcal{X}$, $\mathbf{E}_P[\psi_{x,kt}(W;\theta_0)] = 0$,

$$\sup_{P\in\mathcal{P}} \mathbf{E}_P \left[\sup_{\theta\in B_\Theta(\delta)} \sup_{x\in\mathcal{X}} |\psi_{x,kt}(W;\theta)|^{2+\eta} \right] < \infty.$$

(c) There exist a bounded function V on $\mathcal{X}$ and constants $C, \delta > 0$ and $s_1 \in (1/2, 1]$ such that for each $(x_1, \theta_1) \in \mathcal{X} \times B_\Theta(\delta)$ and for each $\varepsilon > 0$, $t \in \mathcal{T}$,

$$\mathbf{E} \left[\sup_{x\in\mathcal{X}: d_V(x,x_1)\le\varepsilon} \; \sup_{\theta\in B_\Theta(\delta): ||\theta-\theta_1||\le\varepsilon} |\psi_{x,kt}(W;\theta) - \psi_{x_1,kt}(W;\theta_1)|^2 \right] \le C\varepsilon^{2s_1},$$

where $d_V(x, x_1) := |V(x) - V(x_1)|$.

Under these regularity conditions, we can derive the asymptotic null distribution of the test statistic, which is stated in Lemma 8 in the appendix. The limiting distribution is non-pivotal and so the issue of how to conduct inference here remains to be settled. We suggest several solutions to this problem starting with a bootstrap procedure to compute the critical values, which extends the contact set approach of Linton, Song and Whang (2010).

4. Bootstrap Critical Values

The asymptotic distribution of the test statistic depends on the true data generating processes in a complex way. The traditional approach for obtaining critical values is to mimic the asymptotic distribution of the test statistic under the least favorable case (LFC) of the null hypothesis, where the inequalities composing the null hypothesis (2.2) hold with equalities. However, this approach is generally too conservative because the LFC is only a strict subset of the null hypothesis. In a dynamic context such as ours, this approach is even less attractive because the null hypothesis consists of inequality restrictions among a *sequence* of distribution functions over possibly a long time period, and the inequalities might be binding only at a subset of the whole time period $\mathcal{T}$ and/or over a subset (i.e., the *"contact set"*) of the support $\mathcal{X}$.

By exploiting the information on the contact set, we propose a method to compute bootstrap critical values that may yield tests with enhanced power. For the purpose of comparison, we first describe the critical values based on the LFC.

4.1. The Least Favorable Case.

Compute the bootstrap critical values in the following steps:

(1) For each $k = 1, 2$, draw a bootstrap sample $\mathcal{S}_k^* := \{\mathbf{W}_{k,1}^*, \ldots, \mathbf{W}_{k,N_k}^*\}$, where the vectors $\mathbf{W}_{k,i}^* = (W_{k0i}^*, \ldots, W_{kTi}^*)^\intercal \in \mathbb{R}^{T+1}$ for $i = 1, \ldots, N_k$ are independently drawn with replacement from the vectors that comprise the original sample $\mathcal{S}_k := \{\mathbf{W}_{k,1}, \ldots, \mathbf{W}_{k,N_k}\}$, where $\mathbf{W}_{k,i} = (W_{k0i}, \ldots, W_{kTi})^\intercal$ represents a time path of W_{kti} over $t \in \mathcal{T}$ for each $i \in \{1, \ldots, N_k\}$.

(2) Using the bootstrap sample $\mathcal{S}_k^*$, compute $X_{kti}^*(\theta) = \varphi_{kt}(W_i^*, \theta)$, the estimate $\hat{\theta}^*$ and the EDFTs:

$$\bar{F}_k^{(n,m)*}(x,t,\theta) = \frac{1}{N_k}\sum_{i=1}^{N_k}\sum_{s=0}^{t}\frac{a_t(s)(x - X_{ksi}^*(\theta))^{m-1}1(X_{ksi}^*(\theta) \le x)}{(m-1)!}, \tag{4.1}$$

$$\bar{D}^{(n,m)*}(x,t,\hat{\theta}^*) = \bar{F}_1^{(n,m)*}(x,t,\hat{\theta}^*) - \bar{F}_2^{(n,m)*}(x,t,\hat{\theta}^*), \tag{4.2}$$

(3) Compute the bootstrap test static under the LFC:

$$T_{N,LF}^* = r_N^p \int_{\mathcal{X}} \Lambda_p\left(\hat{v}_1^*(x) - \hat{v}_1(x), \ldots, \hat{v}_L^*(x) - \hat{v}_L(x)\right) dx, \tag{4.3}$$

where $\hat{v}_l^*(x)$ denotes $v_l(x)$, $l = 1, \ldots, L$ with $D^{(i,j)}(x,t)$ replaced by $\bar{D}^{(i,j)}(x,t,\hat{\theta}^*)$.

(4) Repeat the steps (1)-(3) above B-times, and compute the $(1-\alpha)$ quantile of the bootstrap distribution of $T_{N,LF}^*$ as the LFC-bootstrap critical value $c_{N,LF}^*(1-\alpha)$.

In Step (1), if W_{1ti} and W_{2ti} are dependent for each $t \in \mathcal{T}$ and $N_1 = N_2 = N$, then we may generate the bootstrap sample $\mathcal{S}^* := \{\mathbf{Z}_1^*, \ldots, \mathbf{Z}_N^*\}$ by drawing vectors $\mathbf{Z}_i^* = (\mathbf{W}_{1,i}^*, \mathbf{W}_{2,i}^*),\ i = 1, \ldots, N$ from the *paired* observations $\mathcal{S} := \{\mathbf{Z}_1, \ldots, \mathbf{Z}_N\}$, where $\mathbf{Z}_i = (\mathbf{W}_{1,i}, \mathbf{W}_{2,i}),\ i = 1, \ldots, N$.[4]

Under regularity conditions, the bootstrap critical values based on the LFC can be shown to yield tests that are asymptotically valid *uniformly* in P. However, they are often too conservative in practice. As an alternative to the LFC-based bootstrap critical value, we propose a bootstrap critical value that can be less conservative but at the expense of introducing an additional tuning parameter.

4.2. **The Contact Set Approach.** To describe out procedure, we first introduce some notation. Define $\mathcal{N}_L := 2^{\mathbb{N}_L} \setminus \emptyset$, i.e., the collection of all the nonempty subsets of $\mathbb{N}_L \equiv \{1, 2, \ldots, L\}$. For any $A \in \mathcal{N}_L$ and $\mathbf{v} = (v_1, \ldots, v_L)^\intercal \in \mathbb{R}^L$, we define $\mathbf{v}_A$ to be $\mathbf{v}$ except that for each $l \in \mathbb{N}_L \backslash A$, the l-th entry of $\mathbf{v}_A$ is zero, and let

$$\Lambda_{A,p}(\mathbf{v}) \equiv \Lambda_p(\mathbf{v}_A). \tag{4.4}$$

[4]Alternative ways to the critical values under the LFC restriction would include bootstrapping from the pooled sample $\{\mathcal{S}_1, \mathcal{S}_1\}$ or using multiplier simulations.

That is, $\Lambda_{A,p}(\mathbf{v})$ is a "censoring" of $\Lambda_p(\mathbf{v})$ outside the index set A. Now, we define the contact sets: for $A \in \mathcal{N}_L$ and for $c_1, c_2 > 0$,

$$B_{N,A}(c_1, c_2) := \left\{ x \in \mathcal{X} : \begin{array}{ll} |r_N v_l(x)| \leq c_1, & \text{for all } l \in A \\ r_N v_l(x) < -c_2, & \text{for all } l \in \mathbb{N}_L / A \end{array} \right\}, \tag{4.5}$$

where $v_l(x)$ is defined in (2.3).

Lemma 3. *Suppose that Assumptions 1, 2 and 3 hold. Suppose further that c_N is a positive sequence such that $c_{N,1}, c_{N,2} \to \infty$ as $N_1, N_2 \to \infty$. Then,*

$$\inf_{P \in \mathcal{P}_0} P \left\{ T_N = r_N^p \sum_{A \in \mathcal{N}_L} \int_{B_{N,A}(c_{N,1}, c_{N,2})} \Lambda_{A,p}(\hat{v}_1(x), \ldots, \hat{v}_L(x)) dx \right\} \to 1,$$

where $\mathcal{P}_0 \subset \mathcal{P}$ is the set of potential distributions of the observed random vector under the null hypothesis $H_0^{(n,m)}$.

Lemma 3 shows that the test statistic T_N is uniformly approximated by an integral with domain restricted to the contact sets $B_{N,A}(c_N)$ in large samples. This result suggests that one may consider a bootstrap procedure that mimics the representation of T_N in Lemma 3.

Let $\{\hat{v}_l^*(x) : l \in \mathbb{N}_L\}$ be the bootstrap counterpart of $\{\hat{v}_l(x) : j \in \mathbb{N}_L\}$ defined in Step (3) above. Then, our bootstrap test statistic is defined as follows:

$$T_N^* = r_N^p \sum_{A \in \mathcal{N}_L} \int_{\hat{B}_{N,A}(\hat{c}_N)} \Lambda_{A,p} \left(\hat{v}_1^*(x) - \hat{v}_1(x), \ldots, \hat{v}_L^*(x) - \hat{v}_L(x) \right) dx,$$

where $\hat{B}_{N,A}(\hat{c}_N)$ denotes the estimated contact set:

$$\hat{B}_{N,A}(\hat{c}_N) := \left\{ x \in \mathcal{X} : \begin{array}{ll} |r_N \hat{v}_l(x)| \leq \hat{c}_N, & \text{for all } l \in A \\ r_N \hat{v}_l(x) < -\hat{c}_N, & \text{for all } l \in \mathbb{N}_L / A \end{array} \right\}, \tag{4.6}$$

where $\hat{c}_N$ is a positive sequence satisfying the following assumption:

Assumption 4. For $\psi_{x,kt}$ in Assumption 3(b), for any $\varepsilon > 0$,

$$P \left\{ \sup_{x \in \mathcal{X}} \left| \sqrt{N_k} \hat{\Gamma}_{kt,P}(x) - \frac{1}{\sqrt{N_k}} \sum_{i=1}^{N_k} \left\{ \psi_{x,kt}(W_i^*; \hat{\theta}) - \frac{1}{N_k} \sum_{i=1}^{N_k} \psi_{x,kt}(W_i; \hat{\theta}) \right\} \right| > \varepsilon \mid \mathcal{W}_N \right\} \to_P 0, \tag{4.7}$$

uniformly in $P \in \mathcal{P}$, where $\mathcal{W}_N$ is the σ-field generated by $\{W_{kti} : i = 1, ..., N_k; k = 1, 2\}$ and $\hat{\Gamma}_{kt,P}(x) = (1/N_k) \sum_{i=1}^{N_k} \left[h_x(\varphi_{ks}(W_i^*, \hat{\theta}^*)) - h_x(\varphi_{ks}(W_i^*, \hat{\theta})) \right]$.

Assumption 5. For each $N \geq 1$, there exist non-stochastic sequences $c_{N,1}, c_{N,2} > 0$ such that $c_{N,1} \leq c_{N,2}$ and

$$\inf_{P \in \mathcal{P}} P\left\{c_{N,1} \leq \hat{c}_N \leq c_{N,2}\right\} \to 1 \text{ and } c_{N,1} + r_N c_{N,2}^{-1} \to \infty.$$

Let $c^*_{N,\alpha}$ be the $(1-\alpha)$-th quantile from the bootstrap distribution of T^*_N. We take

$$c^*_{N,\alpha,\eta} = \max\{c^*_{N,\alpha}, \eta\} \tag{4.8}$$

as our critical value, where $\eta \equiv 10^{-6}$ is a small fixed number.

The following theorem establishes the uniform validity of our bootstrap procedure.

Theorem 4. *Suppose that Assumptions 1 - 5 hold. Then*

$$\limsup_{N_1, N_2 \to \infty} \sup_{P \in \mathcal{P}_0} P\left\{T_N > c^*_{N,\alpha,\eta}\right\} \leq \alpha.$$

Remark 1. The critical value (4.8) is in a different form from that of Linton, Song and Whang (2010) who considered the SD test in a static context (i.e., $T = 0$) without time preference and proposed to use the LFC-based bootstrap critical value $c^*_{N,LF}(1-\alpha)$ when the estimated contact set $\cup_{A \in \mathcal{N}_L} \hat{B}_{N,A}(\hat{c}_N)$ is empty. By taking maximum with arbitrary small positive constant, the critical value $c^*_{N,\alpha,\eta}$ has substantial computational advantage over the latter.

Remark 2. One might question whether the bootstrap test is asymptotically exact, i.e., the inequality in Theorem 4 holds as an equality. In fact, under some additional assumptions, we can show that the test achieves the level α asymptotically, uniformly over a subset $\mathcal{P}_{00} \subset \mathcal{P}_0$, i.e.,

$$\limsup_{N_1, N_2 \to \infty} \sup_{P \in \mathcal{P}_{00}} \left| P\left\{T_N > c^*_{N,\alpha,\eta}\right\} - \alpha \right| = 0.$$

For brevity, however, we do not give the formal result; see Linton, Song and Whang (2010, Theorem 2 (ii)) and Lee, Song and Whang (2018, Theorem 2) for related results.

We now establish consistency of our proposed test:

Theorem 5. *Suppose that Assumptions 1 - 5 hold. Then, under a fixed alternative hypothesis $H_1^{(n,m)}$ such that*

$$\int_{\mathcal{X}} \Lambda_p\left(v_1(x), \ldots, v_L(x)\right) dx > 0,$$

we have, as $N_1, N_2 \to \infty$,

$$P\left\{T_N > c^*_{N,\alpha,\eta}\right\} \to 1.$$

5. Alternative critical values

The contact set approach described above achieves power enhancement using the bootstrap critical values based on explicit estimates of the binding parts of the inequality constraints. In this section, we consider the critical values based on the numerical delta method, which also might have enhanced power properties compared to the traditional LFC-based critical values and have some computational advantages.

The numerical delta method can be described using the general framework of Fang and Santos (2019) and Hong and Li (2018) for conducting inference on directionally differentiable functions. For this purpose, we first introduce the concept of Hadamard directional differentiability of a map between normed spaces:

Definition 6. Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces endowed with norm $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$, respectively. A map $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$ is said to be *Hadamard directionally differentiable at* $\theta \in \mathbb{D}_\phi$ *tangentially to a set* $\mathbb{D}_0 \subseteq \mathbb{D}$, if there exists a continuous map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{N\to\infty} \left\| \frac{\phi(\theta + t_N h_N) - \phi(\theta)}{t_N} - \phi'_\theta(h) \right\|_{\mathbb{E}} = 0, \tag{5.1}$$

for all sequences $\{h_N\} \subset \mathbb{D}$ and $\{t_N\} \subset \mathbb{R}_+$ such that $t_N \downarrow 0$, $h_N \to h \in \mathbb{D}_0$ as $N \to \infty$ and $\theta + t_N h_N \in \mathbb{D}_\phi$ for all N.

We call $\phi'_\theta(h)$ as the *Hadamard directional derivative at* θ *in direction* h. The map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ is possibly nonlinear, but (5.1) implies that it is continuous and homogeneous of degree one (Shapiro (1990)). If ϕ'_θ is linear, then we say that ϕ is *Hadamard differentiable at* θ *tangentially to* $\mathbb{D}_0$.

Fang and Santos (2019, Theorem 2.1) show that if $\hat{\theta}_N$ is an estimator of $\theta_0 \in \mathbb{D}_\phi$ such that $r_N(\hat{\theta}_N - \theta_0) \Rightarrow \mathbb{G}_0$ for some sequence $r_N \uparrow \infty$, where $\mathbb{G}_0$ is a tight process, then $r_N\left[\phi(\hat{\theta}_N) - \phi(\theta_0)\right] \Rightarrow \phi'_{\theta_0}(\mathbb{G}_0)$. Heuristically, this result follows from the definition (5.1) with $t_N = 1/r_N$ which implies

$$r_N\left[\phi(\hat{\theta_N}) - \phi(\theta_0)\right] = \frac{\phi\left\{\theta_0 + (1/r_N)\cdot r_N(\hat{\theta_N} - \theta_0)\right\} - \phi(\theta_0)}{1/r_N} \approx \phi'_{\theta_0}(r_N(\hat{\theta}_N - \theta_0))$$

and the continuous mapping theorem applied to ϕ'_{θ_0}. Based on this result, Fang and Santos (2019, Theorem 3.2) suggest that the limiting distribution $\phi'_{\theta_0}(\mathbb{G}_0)$ can be consistently estimated by $\hat{\phi}'_N(\mathbb{Z}^*_N)$, where $\mathbb{Z}^*_N$ is a consistent estimator (such as the bootstrap) of $\mathbb{G}_0$ and the map $\hat{\phi}'_N$ is a consistent estimator of ϕ'_{θ_0} satisfying their Assumption 4. This method is closely related to the contact set approach described above.

On the other hand, Hong and Li (2018, hereafter HL) suggest a numerical delta method that does not require analytic computation of the directional derivative ϕ'_{θ_0} which can be cumbersome in some applications. They propose estimating $\phi'_{\theta_0}(\mathbb{G}_0)$ by the numerical derivative with step size ϵ_N:

$$\tilde{\phi}'_N(\mathbb{Z}^*_N) := \frac{\phi(\hat{\theta}_N + \epsilon_N \mathbb{Z}^*_N) - \phi(\hat{\theta}_N)}{\epsilon_N}, \tag{5.2}$$

where ϵ_N satisfies $\epsilon_N \to 0$ and $r_N \epsilon_N \to \infty$ as $N \to \infty$. Heuristically, this estimator works because

$$\tilde{\phi}'_N(\mathbb{Z}^*_N) = \frac{\phi\left\{\theta_0 + \epsilon_N\left(\mathbb{Z}^*_N + \frac{\hat{\theta}_N - \theta_0}{\epsilon_N}\right)\right\} - \phi(\theta_0)}{\epsilon_N} - \frac{\phi\left\{\theta_0 + \epsilon_N\left(\frac{\hat{\theta}_N - \theta_0}{\epsilon_N}\right)\right\} - \phi(\theta_0)}{\epsilon_N} \approx \phi'_{\theta_0}(\mathbb{G}_0),$$

where the approximation follows from the definition (5.1) and the results $\frac{\hat{\theta}_N - \theta_0}{\epsilon_N} = \frac{r_N\left(\hat{\theta}_N - \theta_0\right)}{r_N \epsilon_N} \approx 0$ and $\mathbb{Z}^*_N + \frac{\hat{\theta}_N - \theta_0}{\epsilon_N} \approx \mathbb{G}_0$ for N sufficiently large.[5]

When the first order numerical derivative is degenerate, the second (or higher) order numerical delta method can be used (HL and Cheng and Fang (2019)). The *second order Hadamard directional derivative* at $\theta_0 \in \mathbb{D}_\phi$ in the direction h tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$ is defined as $\phi''_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{N \to \infty} \left\| \frac{\phi(\theta_0 + t_N h_N) - \phi(\theta_0) - t_N \phi'_{\theta_0}(h_N)}{\frac{1}{2} t_N} - \phi''_{\theta_0}(h) \right\|_{\mathbb{E}} = 0, \tag{5.3}$$

for all sequences $\{h_N\} \subset \mathbb{D}$ and $\{t_N\} \subset \mathbb{R}_+$ such that $t_N \downarrow 0$, $h_N \to h \in \mathbb{D}_0$ as $N \to \infty$ and $\theta_0 + t_N h_N \in \mathbb{D}_\phi$ for all N. If $\hat{\theta}_N$ satisfies $r_N(\hat{\theta}_N - \theta_0) \Rightarrow \mathbb{G}_0$ for some sequence $r_N \uparrow \infty$, where $\mathbb{G}_0$ is a tight process, then it can be shown (Theorem 4.1 of HL) that

$$r_N^2 \left[\phi(\hat{\theta}_N) - \phi(\theta_0) - \phi'_{\theta_0}(\hat{\theta}_N - \theta_0)\right] \Rightarrow \frac{1}{2} \phi''_{\theta_0}(\mathbb{G}_0).$$

[5] See Theorem 3.1 of HL for a formal result.

If $\phi'_{\theta_0}(h) = 0\ \forall h \in \mathbb{D}_0$, we may approximate $\frac{1}{2}\phi''_{\theta_0}(\mathbb{G}_0)$ by the second order numerical derivative:

$$\frac{1}{2}\tilde{\phi}''_N(\mathbb{Z}^*_N) := \begin{cases} \frac{\phi(\hat{\theta}_N+\epsilon_N\mathbb{Z}^*_N)-\phi(\hat{\theta}_N)}{\epsilon_N^2} & \text{or} \\ \frac{\phi(\hat{\theta}_N+2\epsilon_N\mathbb{Z}^*_N)-2\phi(\hat{\theta}_N+\epsilon_N\mathbb{Z}^*_N)+\phi(\hat{\theta}_N)}{2\epsilon_N^2}, & \end{cases} \tag{5.4}$$

where ϵ_N satisfies $\epsilon_N \to 0$ and $r_N\epsilon_N \to \infty$ as $N \to \infty$.

In our setting, we let

$$\phi(\theta_0) = \int_{\mathcal{X}} \Lambda_p \left(F_1^{(1,1)}(x,T) - F_2^{(1,1)}(x,T), \ldots, F_1^{(n,m)}(x,T-1) - F_2^{(n,m)}(x,T-1)\right) dx,$$

where $\theta_0 = \left(F_1^{(1,1)}(\cdot,T), F_2^{(1,1)}(\cdot,T), \ldots, F_1^{(n,m)}(\cdot,T-1), F_2^{(n,m)}(\cdot,T-1)\right) \in \prod_{l=1}^{2L}\ell^\infty(\mathcal{X})$,[6] $\mathbb{D} = \prod_{l=1}^{2L}\ell^\infty(\mathcal{X})$, $\mathbb{E} = \mathbb{R}$, and $\phi : \prod_{l=1}^{2L}\ell^\infty(\mathcal{X}) \to \mathbb{R}$ is defined by

$$\phi(\theta) = \int_{\mathcal{X}} \Lambda_p \left(\theta_1^{(1)}(x) - \theta_2^{(1)}(x), \ldots, \theta_L^{(1)}(x) - \theta_L^{(2)}(x)\right) dx \tag{5.5}$$

for any $\theta := \left(\theta_1^{(1)}, \theta_1^{(2)}, \ldots, \theta_L^{(1)}, \theta_L^{(2)}\right) \in \prod_{l=1}^{2L}\ell^\infty(\mathcal{X})$. Let

$$\hat{\theta}_N = \left(\bar{F}_1^{(1,1)}(\cdot,T), \bar{F}_2^{(1,1)}(\cdot,T), \ldots, \bar{F}_1^{(n,m)}(\cdot,T-1), \bar{F}_2^{(n,m)}(\cdot,T-1)\right) \tag{5.6}$$

denote the estimator of θ_0 (Equation (3.1)). Then, our test statistic can be written as $T_N = r_N^p\phi(\hat{\theta}_N)$.

Let $\mathcal{N}_L$, $\mathbb{N}_L$, and $\Lambda_{A,p}(\mathbf{v})$ be as defined before. For any $A \in \mathcal{N}_L$, define

$$B_A^0(\theta) := \left\{ x \in \mathcal{X} : \begin{array}{ll} \theta_l^{(1)}(x) = \theta_l^{(2)}(x), & \text{for all } l \in A \\ \theta_l^{(1)}(x) < \theta_l^{(2)}(x), & \text{for all } l \in \mathbb{N}_L/A \end{array} \right\}. \tag{5.7}$$

It is straigthforward to show that that the map ϕ is Hadamard directionally differentiable:

Theorem 7. *Let $\phi(\theta)$ and $\Lambda_{A,p}(\mathbf{v})$ be as defined in (5.5) and (4.4), respectively.*

(i) *When $p = 1$, ϕ is first order Hadamard directionally differentiable at any $\theta \in \prod_{l=1}^{2L}\ell^\infty(\mathcal{X})$ satisfying $\theta_l^{(1)} \le \theta_l^{(2)}$, $l = 1, \ldots, L$, and its derivative is given by: for any $h = (h_1^{(1)}, h_1^{(2)}, \ldots, h_L^{(1)}, h_L^{(2)}) \in \prod_{l=1}^{2L}\ell^\infty(\mathcal{X})$,*

$$\phi'_\theta(h) = \sum_{A\in\mathcal{N}_L} \int_{B_A^0(\theta)} \Lambda_{A,1}(h_1^{(1)}(x) - h_1^{(2)}(x), \ldots, h_L^{(1)}(x) - h_L^{(2)}(x))dx.$$

(ii) *When $p = 2$, ϕ is second order Hadamard directionally differentiable at any $\theta \in \prod_{l=1}^{2L}\ell^\infty(\mathcal{X})$ satisfying $\theta_l^{(1)} \le \theta_l^{(2)}$, $l = 1, \ldots, L$, and its first and second derivatives*

[6] See (2.3) for the ordering of the elements of θ_0.

are given by: for any $h = (h_1^{(1)}, h_1^{(2)}, \ldots, h_L^{(1)}, h_L^{(2)}) \in \prod_{l=1}^{2L} \ell^\infty(\mathcal{X})$, $\phi'_\theta(h) = 0$ *and*

$$\phi''_\theta(h) = \sum_{A \in \mathcal{N}_L} \int_{B_A^0(\theta)} \Lambda_{A,2}(h_1^{(1)}(x) - h_1^{(2)}(x), \ldots, h_L^{(1)}(x) - h_L^{(2)}(x))dx.$$

In our setting, with $\hat{\theta}_N$ given by (5.6), we can establish that $r_N\left(\hat{\theta}_N - \theta_0\right)$ converges weakly to a tight Gaussian process (Lemma 8 in the appendix). Therefore, Theorem 7 suggests that we may approximate the asymptotic null distribution of our test statistic $T_N = r_N^p \phi(\hat{\theta}_N)$ by the numerical derivative estimators $\tilde{\phi}'_N(\mathbb{Z}_N^*)$ (equation (5.2), when $p = 1$) and $\frac{1}{2}\tilde{\phi}''_N(\mathbb{Z}_N^*)$ (equation (5.4), when $p = 2$) with $\mathbb{Z}_N^*$ given by

$$\mathbb{Z}_N^* = \left(r_N\left(\hat{v}_1^*(\cdot) - \hat{v}_1(\cdot)\right), \ldots, r_N\left(\hat{v}_L^*(\cdot) - \hat{v}_L(\cdot)\right)\right),$$

where $\{\hat{v}_l^* : l = 1, \ldots, L\}$ is as defined in the bootstrap procedure (4.3). Let $c_{N,ND}^*(1-\alpha)$ denote the $(1-\alpha)$ quantile of the bootstrap distribution of $\tilde{\phi}'_N(\mathbb{Z}_N^*)$ when $p = 1$ (or $\frac{1}{2}\tilde{\phi}''_N(\mathbb{Z}_N^*)$ when $p = 2$). Then, using Theorems 3.5 and 4.1 of HL, we can show that the test based on the $c_{N,ND}^*(1-\alpha)$ also has (uniformly) correct size under the null hypothesis and is consistent against the fixed alternative.

6. Monte Carlo Experiments

In this section, we conduct Monte Carlo simulations to compare the finite sample performance of the proposed tests. We evaluate the size control of all the procedures we introduced and we evaluate the power improvement of the contact set approach and the numerical delta method. We also compare the finite sample performances of L_1 and L_2 type statistics.

To construct $\hat{c}_N$ for contact-set estimation, we choose

$$\hat{c}_N = C_{cs}(\log\log N) q_{1-\alpha_N}(S_N^*), \tag{6.1}$$

where

$$S_N^* = \max\{\sup_{l,x}(\hat{v}_l^*(x) - \hat{v}_l(x)), \epsilon\sqrt{\log N}\}$$

and $q_{1-\alpha_N}(S_N^*)$ is the $1-\alpha_N$-th quantile of the bootstrap distribution of S_N^* with $\alpha_N = 0.1/\log N$, C_{cs} is a constant and ϵ is a small number.

6.1. **Normal Process.** First, we evaluate a simple stationary normal process. Let $\mathcal{T} \equiv \{0, 1, 2, 3, 4\}$, and define:

$$X_{1,t} = 0.5X_{1,t-1} + Z_t \text{ for } t \in \mathcal{T}$$
$$X_{2,t} \overset{d}{=} X_{1,t} \text{ for } t \in \mathcal{T} \setminus \tau$$
$$X_{2,\tau} \overset{d}{=} \sigma X_{1,\tau},$$

where $X_{1,0}, Z_t \sim N(0,1)$ for $t \in \{1,2,3,4\}$, $\sigma \in \{1, 0.5, 2\}$. The hypotheses of interest are:

$$H_0^{(n,m)} : X_1 \succeq_{nTmSD} X_2 \text{ for } n, m = 1, 2. \tag{6.2}$$

From this DGP, X_1 and X_2 might be distributionally different only at $t = \tau$. If $\sigma = 1$, the two processes are stochastically equivalent, which corresponds to the least favorable case of the null hypothesis. We expect the rejection ratio of the test is near the significance level $\alpha = 0.05$. The case $\sigma = 0.5$ results in crossing of two distributions at $t = \tau$, so we expect rejection of $H_0^{(1,1)}$ and $H_0^{(2,1)}$. In addition, if $\sigma = 0.5$, risk-averse agents would prefer X_2 to X_1, because the two prospects have the same mean over time but the standard deviation of X_2 is smaller. Thus, we also expect the rejection of $H_0^{(1,2)}$ and $H_0^{(2,2)}$. If $\sigma = 2$, there is also crossing of two distributions at $t = \tau$, so we expect rejection of $H_0^{(1,1)}$ and $H_0^{(2,1)}$. However, setting $\sigma = 2$ makes X_1 have a smaller standard deviation than X_2, so $H_0^{(1,2)}$ and $H_0^{(2,2)}$ are true.

Regarding the simulations, we used the grid of size 100 that is equally spaced on the range of pooled EDFTs. We conducted 200 bootstrap resamples and 1000 simulations. For calculating the test statistics, the trapezoidal numerical integration was employed. For tuning parameters, $\eta = 10^{-6}$, $C_{cs} = 0.5$ were selected for both $p = 1$ and $p = 2$ statistics. For the numerical delta method, we choose $\epsilon_N = r_N^{-1/8}$ for $p = 1$ and $\epsilon_N = r_N^{-1/128}$ for $p = 2$. We approximated $\tilde{\phi}'_N(\mathbb{Z}_N^*) \approx \frac{\phi(\hat{\theta}_N + \epsilon_N \mathbb{Z}_N^*) - \phi(\hat{\theta}_N)}{\epsilon_N}$ for L_1-type statistic and $\frac{1}{2}\tilde{\phi}''_N(\mathbb{Z}_N^*) \approx \frac{\phi(\hat{\theta}_N + \epsilon_N \mathbb{Z}_N^*) - \phi(\hat{\theta}_N)}{\epsilon_N^2}$ for L_2-type statistic . Although we used both max and sum type Λ_p, because the two statistics show similar results, we only report results using max type statistics. The sample size N is 500.

We first show the case $\tau = 0$ for comparing the LFC, contact-set approach and the numerical delta method. In Table 1, the simulation results are shown. First, all the LFC, the contact set approach, and the numerical delta method show proper size control as

Table 1. The Rejection Ratio: Normal Process, $\tau = 0$

		$p=1$			$p=2$		
σ	Order	LFC	Contact	NDM	LFC	Contact	NDM
1	(1,1)	0.048	0.05	0.051	0.051	0.051	0.056
	(1,2)	0.056	0.056	0.058	0.053	0.053	0.061
	(2,1)	0.054	0.055	0.056	0.051	0.052	0.059
	(2,2)	0.058	0.058	0.058	0.059	0.059	0.067
0.5	(1,1)	0.116	0.236	0.137	0.34	0.388	0.394
	(1,2)	0.113	0.115	0.113	0.097	0.097	0.109
	(2,1)	0.399	0.686	0.472	0.716	0.822	0.775
	(2,2)	0.17	0.193	0.171	0.164	0.174	0.184
2	(1,1)	0.534	0.66	0.656	0.683	0.752	0.77
	(1,2)	0.009	0.028	0.011	0.026	0.031	0.029
	(2,1)	0.895	0.978	0.946	0.952	0.982	0.964
	(2,2)	0.005	0.025	0.005	0.021	0.027	0.025

Notes The number in each cell of LFC, Contact and NDM columns means the rejection ratio of the simulation using the LFC-based approach, the contact set approach, and numerical delta method to estimate critical value, respectively. For tuning parameters, $\eta = 10^{-6}$, $C_{cs} = 0.5$ were selected for both $p = 1$ and $p = 2$ statistics. For numerical delta method, $\epsilon_N = r_N^{-1/8}$ for $p = 1$ and $\epsilon_N = r_N^{-1/128}$ for $p = 2$ were specified. We used the max-type Λ_p. The test used the grid of size 100 that equally spaced on the range of pooled EDFTs. We conducted bootstrapping 200 times and 1000 simulations. The sample size N is 500.

expected for both $p = 1$ and $p = 2$. Second, power improvement is remarkable when two distributions cross each other. This can be seen at the case $\sigma = 0.5, 2$ for $(1, 1)$ and $(2, 1)$ orders.

We turn our attention to see whether a time order affects the testing result. Note that the second time order preference distinguishes two periods near to the present more than two periods far from the present. For example, if an economic agent has the time preference contained in the second time order preference set, he differentiates today and tomorrow more than 100 days and 101 days after today. Therefore, we can intuitively guess the closer τ is to 0, the more testing results are sensitive to σ. Table 2 shows the result of using the contact-set approach. Remarkably, as time order changes from first to second, the impact of τ gets significant. In columns of TSD order $(2, 1)$ and $(2, 2)$, we can see that the rejection ratio gets bigger and smaller in order, depending on the hypotheses.

Table 2. The Rejection Ratio: Normal Process, Varying τ

		$p=1$				$p=2$			
σ	τ	$(1,1)$	$(2,1)$	$(1,2)$	$(2,2)$	$(1,1)$	$(2,1)$	$(1,2)$	$(2,2)$
0.5	0	0.236	0.686	0.115	0.193	0.388	0.822	0.097	0.174
	2	0.324	0.22	0.158	0.141	0.424	0.324	0.137	0.124
	4	0.31	0.068	0.126	0.072	0.367	0.079	0.107	0.062
2	0	0.66	0.978	0.028	0.025	0.752	0.982	0.031	0.027
	2	0.735	0.619	0.029	0.03	0.726	0.561	0.036	0.033
	4	0.711	0.135	0.04	0.044	0.645	0.103	0.044	0.046

Notes The number in each cell of $(1,1)$, $(1,2)$ $(2,1)$ and $(2,2)$ columns means the rejection ratio of the simulation whose hypotheses are $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$ time stochastic orders respectively. To estimate critical value, we use the contact-set approach. For tuning parameters, $\eta = 10^{-6}$, $C_{cs} = 0.5$ were selected for both $p=1$ and $p=2$ statistics. We used the max-type Λ_p. The test used the grid of size 100 that equally spaced on the range of pooled EDFTs. We conducted 200 times of bootstrapping for all 1000 number of simulations. The sample size N is 500.

6.2. **Mean-Shifting Uniform Distributions.** To explore the relationship between the time order and testing results, we examine simple one-time mean-shifting uniform distributions. We make a shift only at one time period to see how our test is affected by the time period when divergence occurs. Let $\mathcal{T} = \{0,1,2,3,4\}$, and define:

$$X_{1,t} = 0.5X_{1,t-1} + u_t - 1/2 \text{ for } t \in \mathcal{T}$$

$$X_{2,t} \overset{d}{=} X_{1,t} \text{ for } t \in \mathcal{T} \setminus \tau$$

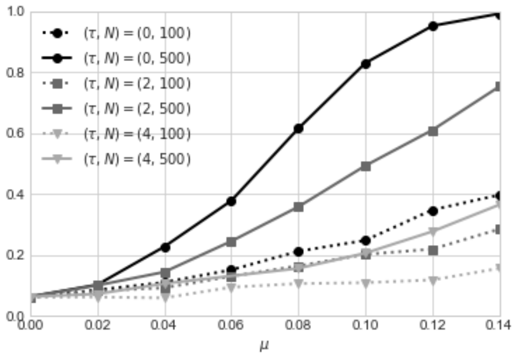
$$X_{2,\tau} \overset{d}{=} X_{1,\tau} + \mu,$$

where $u_t \sim U[0,1]$ for $t \in \mathcal{T}$, $X_{1,-1} \equiv 0$, and $\tau \in \{0,2,4\}$. We vary μ in the range [-0.10,0.14] .

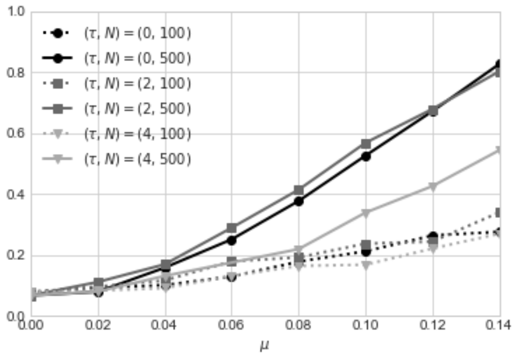
The hypotheses are the same as (6.2). Adding $\mu > 0$ shifts X_1 to become interior of X_2, so the above $H_0^{(n,m)}$ is false for all $n, m = 1, 2$. That is, the expected NPV of X_2 is bigger than the expected NPV of X_1 Thus, we expect the rejection of all hypotheses. In contrast, if we make $\mu < 0$, X_2 is interior of X_1, so the hypotheses are true. Simulation was done in the same setting in Section 6.1. except we compare two sample sizes $N = 100, 500$.

In this design, by varying μ and τ, we show how the magnitude of shifting and time order jointly affect the power. In Figure 6.1, the test results are reported. We only show the results of the contact-set approach. The other two methods show qualitatively similar results. In addition, we did simulations using both the max and sum type Λ_p for $p = 1$ and $p = 2$ cases. Because all specifications show similar results, we only report simulation results from utilizing the sum type Λ_p with $p = 2$.

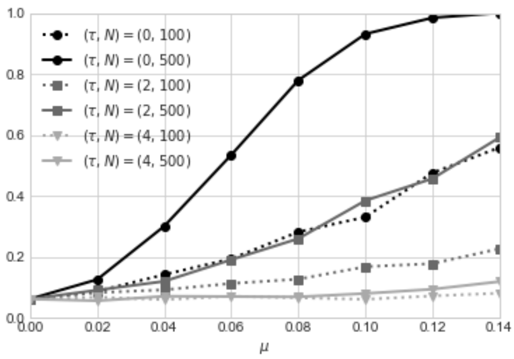
Figure 6.1. Simulation: Time order and Divergence Period (Power)



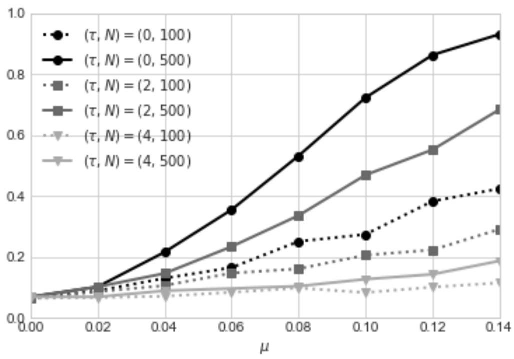
A. (1,1) Order



B. (1,2) Order



C. (2,1) Order

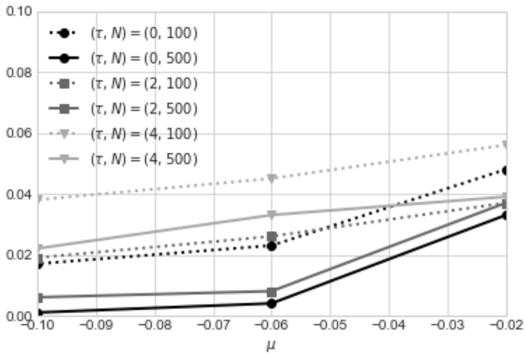


D. (2,2) Order

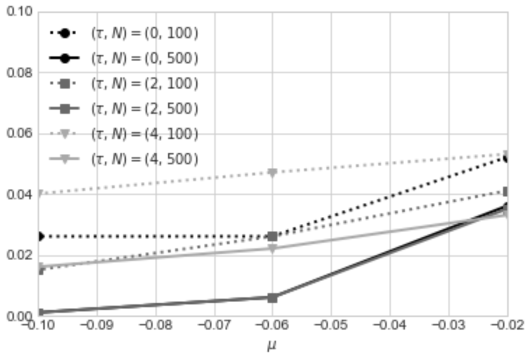
Notes This figure shows the rejection ratio of each specified (τ, N). Panel A. B. C. and D. are the results of $(1, 1)$, $(1, 2)$, $(2, 1)$, and $(2, 2)$ time stochastic order hypotheses, respectively. The solid line and the dotted line represent results of the sample size $N = 500$ and $N = 100$, respectively. The round-, square- and down-triangle symbol represent $\tau = 0$, $\tau = 2$, and $\tau = 4$, respectively. We used the grid of size 100 that is equally spaced on the range of pooled EDFTs. We conducted 200 times of bootstrapping and 1000 number of simulations. For tuning parameters, $\eta = 10^{-6}$, $C_{cs} = 0.5$ were selected. The sum type Λ_p with $p = 2$ was specified.

In Figure 6.1, in the second time order testing, test results are more affected by τ. By comparing results of Panel A and C, and Panel B and D, we can see that the gap between $\tau = 0$ and $\tau = 4$ lines becomes larger as time order goes up from the first to the second for both sample sizes. The gap between $\tau = 0$ and $\tau = 4$ lines also becomes more salient.

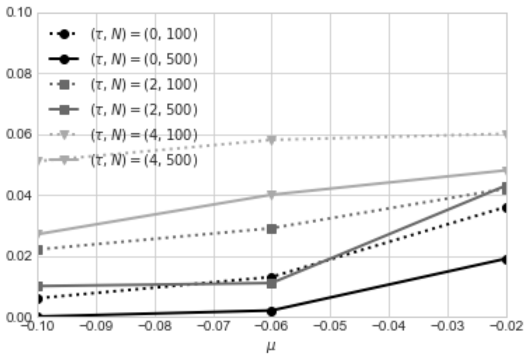
FIGURE 6.2. Simulation: Time order and Divergence Period (Size)



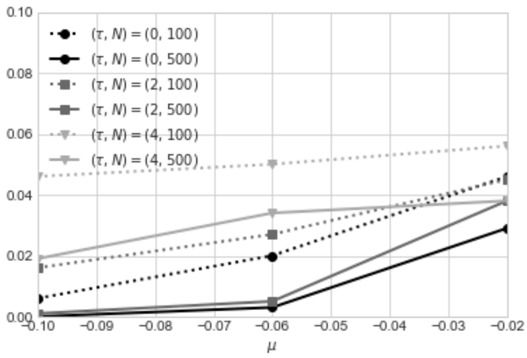
A. (1,1) Order



B. (1,2) Order



C. (2,1) Order



D. (2,2) Order

Notes This figure shows the rejection ratio of each specified (τ, N). Panel A. B. C. and D. are the results of $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$ time stochastic order hypotheses, respectively. The solid line and the dotted line represent results of the sample size $N = 500$ and $N = 100$, respectively. The round-, square- and down-triangle symbol represent $\tau = 0$, $\tau = 2$, and $\tau = 4$, respectively. We used the grid of size 100 that is equally spaced on the range of pooled EDFTs. We conducted 200 times of bootstrapping and 1000 number of simulations. For tuning parameters, $\eta = 10^{-6}$, $C_{cs} = 0.5$ were selected. The sum type Λ_p with $p = 2$ was specified.

In Figure 6.2., we show test results when our hypotheses are true. Our test well controls the size under or near to significance level $\alpha = 0.05$. Again, the testing results are more sensitive to τ when the time order is the second.

7. Empirical Study

7.1. **Testing the Welfare Improvement of a Microfinance Program.** Does development policy result in the improvement of the long-term welfare? This is one of the key questions addressed in the economics literature. In this section, we illustrate how our test can be applied for policy evaluation in a dynamic context. Our test provides a general dynamic welfare comparison for evidence-based policy-decisionmaking.

We evaluate the impact of Thailand's Million Baht Village Fund Program on dynamic welfare; this project was evaluated in Kabsoki and Townsend (2011) using a structural model and in Kabsoki and Townsend (2012) by a reduced-form analysis. The relationship between development and credit constraints is important in terms of welfare as many economists have discussed (e.g. Banerjee and Newman 1993; Galor and Zeira 1993; Banerjee and Duflo 2014). We use the replicate data of Kabsoki and Townsend (2012) that is publicly available on the *American Economic Journal: Applied Economics* website.[7] This data is from the Townsend Thai Survey (Townsend et al. 1997), which contains 11 years (1997-2007) of longitudinal households records from 64 rural Thai villages.[8] Our test is suitable for such a large field panel data structure.

To compare the impact of different amount of funds per household, we split households into two groups based on the village median size in 2001. Because each village received the same amount of funding, regardless of the size, a smaller village received more funds per household.[9] Therefore, we compare the dynamic utility of two household groups: households in a small village and in a big village. For the utility measure, we use annual total consumption (TC) and net income (NI).

We control for systematic differences induced by factors other than the fund program by a linear regression model. Specifically, we follow the regression equation (6) in Kaboski and Townsend (2012) with a slight modification

$$y_{n,t,k} = \sum_{i=1}^{I} \beta_{i,k} Z_{i,n,t} + \phi_{t,k} + \epsilon_{n,t,k}, \tag{7.1}$$

where subscript n means a household, t means a time period, and k ($\in \{TC, NI\}$) indicates a welfare measure, while $y_{n,t,k}$ is either total consumption or net income, and $Z_{i,n,t}$

[7] `https://doi.org/10.1257/app.4.2.98`

[8] For more detail, refer to Townsend et al. (1997) and Kaboski and Townsend (2012).

[9] See Kaboski and Townsend (2011, 2012) about the village size and exogeneity of the funding program.

contains household control variables including: the number of adult males, the number of adult females, the number of children, a dummy for male head of household, the age of household head, the age of head squared, and the years of schooling of the household head. ϕ_t is a yearly time fixed-effect. Compared with the original model, our specification excludes household fixed effects and an intercept term. This is because our test allows general time dependence structure of each household, so a fixed effect term is unnecessary. Let $Y_{v,k}$ be a k-variable's stochastic process of v group($v \in \{Small, Big\}, k \in \{TC, NI\}$) from 2002 to 2007 and $\theta = (\beta_1, \ldots, \beta_I, \phi_{2002}, \ldots, \phi_{2007})^\intercal$. Also, let $\boldsymbol{\epsilon}(\hat{\theta})_{v,k}$ be a residual k stochastic process of group v.

Figure 7.1 shows descriptive analysis. In Figure 7.1 A, regarding the small group, there is an increase in mean total consumption in 2002 that leads to a persistent gap for 4 years. However, the trend of median total consumption does not show any significant change. In Figure 7.1 B, both mean and median trends of each group's net income cross each other many times. There is an increase in net income only in 2003 but it does not lead to a persistent gap. In Figure 7.1 C and D, we show standard deviations. The small village group's standard deviation of total consumption grows a lot after 2002, which is not the case for the net income measure. This figure implies the necessity of TSD testing for two reasons. First, huge differences between the mean and the median levels of both total consumption and net income indicate there are outliers who consume and earn way more than other households. Comparing only a mean level is vulnerable to such outliers and policy evaluation only using the mean level cannot suggest an overall picture of dynamic welfare improvement. Thus, taking distributions as a whole is necessary. Second, even though there is a gap between the mean total consumption of each group, the standard deviation also shows a large gap. Thus, standard mean-variance analysis also cannot determine whether there was more welfare improvement for the small village group.

The hypotheses are as follows:

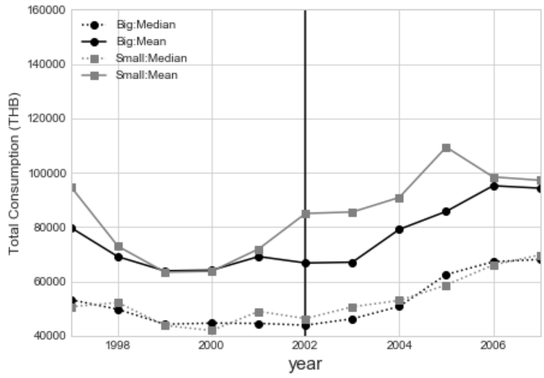
$$H_{0,1}^{(n,m)} : Y_{Small,k} \succeq_{nTmSD} Y_{Big,k} \tag{7.2}$$

$$H_{0,2}^{(n,m)} : Y_{Big,k} \succeq_{nTmSD} Y_{Small,k} \tag{7.3}$$

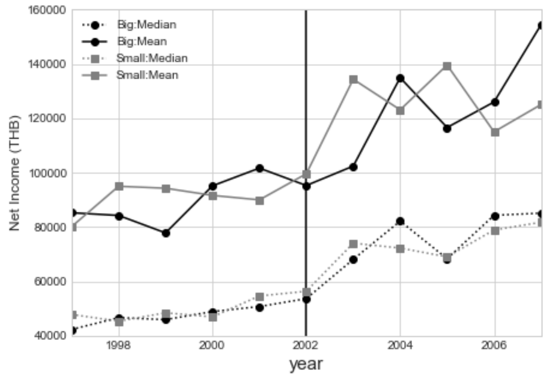
$$H_{0,3}^{(n,m)} : \boldsymbol{\epsilon}(\hat{\theta})_{Small,k} \succeq_{nTmSD} \boldsymbol{\epsilon}(\hat{\theta})_{Big,k} \tag{7.4}$$

$$H_{0,4}^{(n,m)} : \boldsymbol{\epsilon}(\hat{\theta})_{Big,k} \succeq_{nTmSD} \boldsymbol{\epsilon}(\hat{\theta})_{Small,k} \tag{7.5}$$

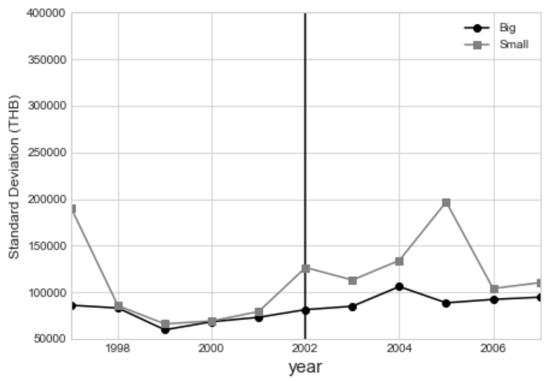
FIGURE 7.1. Mean, quantiles and standard deviation of total consumption and net income



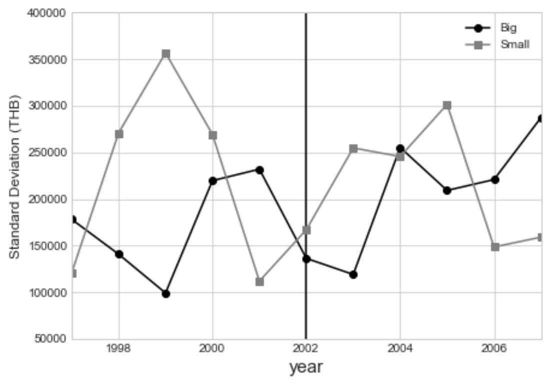
A. Total Consumption: Mean and Quantiles



B. Net Income: Mean and Quantiles



C. Total Consumption: Standard Deviation



D. Net Income: Standard Deviation

Notes This figure shows the trend of yearly mean and quantiles of each small and big-size village group from 1997 to 2007 year. The cross-sectional sample size is 344 and 328 for small- and big-size village group, respectively. In Figure A and B, the solid lines indicate mean, the dotted lines shows median. The line in C and D indicate the trend of standard deviation of total consumption and net income, respectively. The black vertical line indicates year 2002 when the Million Baht Village Fund Program initiated. We use replicate data by Kaboski and Townsend (2012)

where $n, m = 1, 2$ and $k = TC, NI$. $H_{1,i}^{(n,m)}$ is the negation of $H_{0,i}^{(n,m)}$ for $i = 1, 2, 3, 4$.

Intuitively, a household with more available funds has fewer credit constraints, so can improve their dynamic utility through maximization. Based on this simple framework, we expect rejection of (7.5) if the econometric model properly controls for other factors. In addition, as a falsification test, we apply the TSD test to data from 1997 to 2001. If there is no TSD relation before the program but there are TSD relations after the program was in place, it would be reasonable to argue that there is a general welfare improvement due to the fund program.

Table 3. P-Values: Testing the Welfare Improvement

	Panel A. Before the Fund Program (1997-2001)							
	Total Consumption				Net Income			
$H_0^{(n,m)}$	(1,1)	(1,2)	(2,1)	(2,2)	(1,1)	(1,2)	(2,1)	(2,2)
$Y_{Small} \succeq Y_{Big}$	0.72	1	0.785	1	0.64	0.57	0.64	0.6
$Y_{Big} \succeq Y_{Small}$	0.305	0.245	0.23	0.165	0.705	0.51	0.68	0.505
$\boldsymbol{\epsilon}_{Small} \succeq \boldsymbol{\epsilon}_{Big}$	0.88	0.78	0.9	0.81	1	0.995	0.995	0.985
$\boldsymbol{\epsilon}_{Big} \succeq \boldsymbol{\epsilon}_{Small}$	0.84	0.255	0.75	0.16	0.7	0.19	0.48	0.165
	Panel B. After the Fund Program (2002-2007)							
	Total Consumption				Net Income			
$H_0^{(n,m)}$	(1,1)	(1,2)	(2,1)	(2,2)	(1,1)	(1,2)	(2,1)	(2,2)
$Y_{Small} \succeq Y_{Big}$	0.92	1	0.89	1	0.645	0.55	0.785	0.58
$Y_{Big} \succeq Y_{Small}$	0.06	0.02	0.035	0.005	0.54	0.265	0.545	0.255
$\boldsymbol{\epsilon}_{Small} \succeq \boldsymbol{\epsilon}_{Big}$	1	0.985	1	0.955	0.995	0.915	1	0.92
$\boldsymbol{\epsilon}_{Big} \succeq \boldsymbol{\epsilon}_{Small}$	0.175	0.02	0.07	0.005	0.375	0.015	0.27	0.015

Notes We test hypotheses (7.2) to (7.5) using panel data both before (1997-2001) and after the fund program (2002-2007) The cross-sectional sample size is 344 and 328 for small- and big-size village group, respectively. We used the contact-set approach. $\eta = 10^{-6}$ and $C_{cs} = 0.5$ were specified as recommended based on simulation results. We used the max-type Λ_p and $p = 1$. The test used the grid of size 100 that equally spaced on the range of pooled EDFTs. We conducted bootstrapping 200 times. $p = 2$ case also reveals qualitatively similar results.

We show the test results in Table 3. For the significance level $\alpha = 0.05$, every hypothesis is not rejected when it comes to the period before the fund program. Thus, two prospects stochastically yield the same expected NPV under general classes of utility and time-discounting functions, corresponding to specified time and stochastic orders. In Panel B, there is a difference between two different welfare measures. First, the hypotheses (7.3) for (1,2), (2,1) and (2,2) TSD order using total consumption are rejected, but hypotheses (7.2) are not rejected for every TSD order. In contrast, both hypotheses (7.2) and (7.3) using net income are not rejected. Interestingly, after controlling for systematic difference, test results become similar for both measurements; we rejected (7.5) for (2,1) and (2,2) TSD order in both cases when using total consumption and net income. Therefore, our testing provides a general consensus on the welfare improvement impact of the fund program conditionally on regional household characteristics.

8. Conclusion

This paper proposed L_p integrated type tests for TSD and attained their asymptotic distributions under standard panel data sampling scheme. We applied a path-wise bootstrap procedure to allow individual time series dependence. We described three methods for attaining critical values, the LFC-based approach, contact-set approach and numerical delta method which all are asymptotically valid. Monte-Carlo simulation results were shown to verify finite sample properties of the test. The relationship between the time order and period when two prospects were different was well illustrated in simulation results. We applied our testing to evaluate the dynamic welfare improvement of a microfinance program in Thailand. The TSD hypothesis implies strong restrictions on the data. Therefore, if the null hypothesis is not rejected, the implied ordering is very convincing.

Because the statistical test for TSD had not yet been developed properly, this paper would take a role as a stepping stone for further tests for TSD. In addition, since empirical analysis to compare dynamic distributions over time as a whole has been rare, this paper sheds light on a distinctive way of implementing empirical work. In particular, it would be straightforward to extend our testing to dynamic counterfactual distributions of different policy scenarios based on certain estimated relations.

9. Appendix

Let $\nu_{kt}(\cdot)$ is a Gaussian process on $\mathcal{X}$ with mean zero and covariance kernel given by

$$C_{kt}(x_1, x_2) = Cov\left(V_{x_1,kt}(W_i;\theta_0), V_{x_2,kt}(W_i;\theta_0)\right),$$

where

$$V_{x,kt}(W_i;\theta) = \sum_{s=0}^{t} a_t(s)\left[h_x(\varphi_{ks}(W_i,\theta)) + \psi_{x,ks}(W_i;\theta)\right].$$

Lemma 8. *Suppose that Assumptions 1, 2 and 3 hold. Then, for all $k \in \{1,2\}$, $n, m \in \mathbb{Z}^+$ and $t \in \mathcal{T}$, we have*

(i) $\sqrt{N_k}\left(\bar{F}_k^{(n,m)}(x,t,\hat{\theta}) - F_k^{(n,m)}(x,t;\theta_0)\right) = \eta_{N,kt}(x) + o_{\mathcal{P}}(1)$, *uniformly in $x \in B(\delta_N)$, where*

$$\eta_{N,kt}(x) = \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\left\{V_{x,kt}(W_i;\theta_0) - E\left[V_{x,kt}(W_i;\theta_0)\right]\right\}.$$

(ii) $r_N\left(\bar{D}^{(n,m)}(\cdot,t,\hat{\theta})-D^{(n,m)}(\cdot,t;\theta_0)\right)\Rightarrow\sqrt{1-\lambda}\nu_{1t}(\cdot)-\sqrt{\lambda}\nu_{2t}(\cdot)$ *in* $l_\infty(\mathcal{X})$ *uniformly in* $P\in\mathcal{P}$, *where* ν_{1t} *and* ν_{2t} *are Gaussian processes on* $\mathcal{X}$ *with mean zero and covariance kernels given by* $C_{1t}(\cdot,\cdot)$ *and* $C_{2t}(\cdot,\cdot)$, *respectively.*

Proof of Lemma 8: (i) By rearranging terms, we write

$$\begin{aligned}&\sqrt{N_k}\left(\bar{F}_k^{(n,m)}(x,t,\hat{\theta})-F_k^{(n,m)}(x,t;\theta_0)\right)\\&=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t}a_t(s)\left\{h_x(\varphi_{ks}(W_i,\hat{\theta}))-\mathbf{E}\left(h_x(\varphi_{ks}(W_i,\theta_0))\right)\right\}\\&=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t}a_t(s)\left\{h_x(\varphi_{ks}(W_i,\theta_0))-\mathbf{E}\left(h_x(\varphi_{ks}(W_i,\theta_0))\right)\right\}\\&+\sqrt{N_k}\sum_{s=0}^{t}a_t(s)\Gamma_{ks,P}(x)^\intercal[\hat{\theta}-\theta_0],\\&+\zeta_{1N}+\zeta_{2N},\end{aligned}$$

where

$$\begin{aligned}\zeta_{1N}&=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t}a_t(s)\left[h_x(\varphi_{ks}(W_i,\hat{\theta}))-h_x(\varphi_{ks}(W_i,\theta_0))\right]\\&-\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t}a_t(s)\left[\mathbf{E}\left(h_x(\varphi_{ks}(W_i,\hat{\theta}))\right)-\mathbf{E}\left(h_x(\varphi_{ks}(W_i,\theta_0))\right)\right],\\\zeta_{2N}&=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t}a_t(s)\left[\mathbf{E}\left(h_x(\varphi_{ks}(W_i,\hat{\theta}))\right)-\mathbf{E}\left(h_x(\varphi_{ks}(W_i,\theta_0))\right)\right]\\&-\sqrt{N_k}\sum_{s=0}^{t}a_t(s)\Gamma_{ks,P}(x)^\intercal[\hat{\theta}-\theta_0].\end{aligned}$$

By Assumption 3(b), we have $\zeta_{2N}=o_{\mathcal{P}}(1)$. Let

$$\mathcal{H}=\left\{\sum_{s=0}^{t}a_t(s)\left[h_x(\varphi_{ks}(\cdot,\theta))-h_x(\varphi_{ks}(\cdot,\theta))\right]:(x,\theta)\in\mathcal{X}\times B_\Theta(\delta_N)\right\}.$$

For any decreasing sequence $\delta_N\to 0$, the bracketing entropy of this class at $\epsilon\in(0,1]$ is bounded by $C\epsilon^{-d\lambda/s_2}$ by Lemma B2 of Linton, Song and Whang (2010). Therefore, using the fact $d\lambda/s_2<2$ and the maximal inequality, we have $\zeta_{1N}=o_{\mathcal{P}}(1)$. Now, Assumption 1(d) gives the desired result of Lemma 8(i).

(ii) Let $\mathcal{F}=\{h+\psi:(h,\psi)\in\mathcal{H}_0\times\Psi_0\}$, where $\mathcal{H}_0=\left\{\sum_{s=0}^{t}a_t(s)h_x(\varphi_{ks}(\cdot,\theta_0)):x\in\mathcal{X}\right\}$ and $\Psi_0=\left\{\sum_{s=0}^{t}a_t(s)\psi_{x,ks}(\cdot;\theta_0):x\in\mathcal{X}\right\}$. Using Assumption 3(c) and by Lemmas B1 and

B2 of Linton, Song and Whang (2010), we can show that $\sup_{P\in\mathcal{P}} \log N_{[]}\left(\epsilon, \mathcal{F}, L_2(P)\right) < C\log\epsilon$. By Theorem 2.3 of Sheehy and Wellner (1991), $\mathcal{F}$ is a uniform Donsker class. Therefore, using Assumption 1(b), we have the desired weak convergence result of Lemma 8(ii). □

Lemma 9. *Suppose that Assumptions 1, 2, 3 and 4 hold. Then, for all $k \in \{1,2\}$, $n, m \in \mathbb{Z}^+$ and $t \in \mathcal{T}$, we have*

(i) $\sqrt{N_k}\left(\bar{F}_k^{(n,m)*}(x,t,\hat{\theta}^*) - \bar{F}_k^{(n,m)}(x,t;\hat{\theta})\right) = \nu^*_{N,kt}(x,\hat{\theta}) + o_{\mathcal{P}^*}(1)$, *in P uniformly in $P\in\mathcal{P}$, where*

$$\nu^*_{N,kt}(x,\theta) = \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\left\{V_{x,kt}(W_i^*;\theta) - \frac{1}{N_k}\sum_{i=1}^{N_k}\left[V_{x,kt}(W_i;\theta)\right]\right\}.$$

(ii) $r_N\left(\bar{D}^{(n,m)*}(\cdot,t,\hat{\theta}^*) - \bar{D}^{(n,m)}(\cdot,t;\hat{\theta})\right) \Rightarrow \sqrt{1-\lambda}\nu_{1t}(\cdot) - \sqrt{\lambda}\nu_{2t}(\cdot)$ *in $l_\infty(\mathcal{X})$ conditional on $\mathcal{W}_N$ in P uniformly in $P\in\mathcal{P}$, where ν_{1t} and ν_{2t} are Gaussian processes on $\mathcal{X}$ with mean zero and covariance kernels given by $C_{1t}(\cdot,\cdot)$ and $C_{2t}(\cdot,\cdot)$, respectively.*

Proof of Lemma 9: (i) Write

$$\begin{aligned}
&\sqrt{N_k}\left(\bar{F}_k^{(n,m)*}(x,t,\hat{\theta}^*) - \bar{F}_k^{(n,m)}(x,t;\hat{\theta})\right)\\
&= \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t} a_t(s)\left[h_x(\varphi_{ks}(W_i^*,\hat{\theta}^*)) - h_x(\varphi_{ks}(W_i,\hat{\theta}))\right]\\
&= \sum_{s=0}^{t} a_t(s)\hat{\Gamma}_{ks,P}(x) + \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\sum_{s=0}^{t} a_t(s)\left[h_x(\varphi_{ks}(W_i^*,\hat{\theta})) - h_x(\varphi_{ks}(W_i,\hat{\theta}))\right],
\end{aligned}$$

where $\hat{\Gamma}_{ks,P}(x)$ is as defined in Assumption 4. Now the desired result holds because

$$\hat{\Gamma}_{ks,P}(x) = \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N_k}\left\{\psi_{x,ks}(W_i^*;\hat{\theta}) - \frac{1}{N_k}\sum_{i=1}^{N_k}\psi_{x,ks}(W_i;\hat{\theta})\right\} + o_{P^*}(1)$$

by Assumption 4.

(ii) Using Linton, Song and Whang (2010, Proof of Lemma B3), the class of functions $H_\delta = \{V_{x,kt}(\cdot;\theta) : (x,\theta)\in\mathcal{X}\times B_\Theta(\delta)\}$ is a bootstrap uniform Donsker class. Therefore, we can show that

$$\sup_{(x,\theta)\in\mathcal{X}\times B_\Theta(\delta)} \left|\nu^*_{N,kt}(x,\theta) - \nu^*_{N,kt}(x,\theta_0)\right| = o_{P^*}(1).$$

Now, the bootstrap uniform CLT applied to $\nu^*_{N,kt}(x,\theta_0)$ gives the desired result. □

Proof of Lemma 3: Define

$$\hat{\mathbf{s}}(x) = [r_N\{\hat{v}_l(x) - v_l(x)\}]_{l\in\mathbb{N}_L},$$
$$\hat{\mathbf{u}}(x) = [r_N\hat{v}_l(x)]_{l\in\mathbb{N}_L} \text{ and } \mathbf{u}(x) = [r_N v_l(x)]_{l\in\mathbb{N}_L}.$$

Let $B_N(c_N) := \cup_{A\in\mathcal{N}_L} B_{N,A}(c_{N,1}, c_{N,2})$. Then it suffices to show the following results:

$$\inf_{P\in\mathcal{P}_0} P\left\{\int_{\mathcal{X}\setminus B_N(c_N)} \Lambda_p(\hat{\mathbf{u}}(x))dx = 0\right\} \to 1, \tag{9.1}$$

$$\inf_{P\in\mathcal{P}_0} P\left\{\int_{B_N(c_N)} \{\Lambda_p(\hat{\mathbf{u}}(x)) - \Lambda_{A,p}(\hat{\mathbf{u}}(x))\}\, dx = 0\right\} \to 1 \tag{9.2}$$

for each $A \in \mathbb{N}_L$, as $N_1, N_2 \to \infty$.

We first establish (9.1). Observe that, whenever $x \in \mathcal{X} \setminus B_N(c_N)$, we have $r_N v_l(x) \leq -c_{N,1}\, \forall l \in \mathbb{N}_L$ under the null hypothesis. Therefore, for all $l \in \mathbb{N}_L$, we have

$$\begin{aligned}\int_{\mathcal{X}\setminus B_N(c_N)} \Lambda_p(\hat{\mathbf{u}}(x))dx &= \int_{\mathcal{X}\setminus B_N(c_N)} \Lambda_p(\hat{\mathbf{s}}(x) + \mathbf{u}(x))dx \\ &\leq \int_{\mathcal{X}\setminus B_N(c_{N,1})} \Lambda_p(\hat{\mathbf{s}}(x) - c_{N,1}\mathbf{1}_L)dx \\ &\leq L^{p/2}\left(\sum_{l=1}^{L}\left[r_N \sup_{x\in\mathcal{X}} |\hat{v}_l(x) - v_l(x)| - c_{N,1}\right]_+^2\right)^{p/2},\end{aligned} \tag{9.3}$$

where $\mathbf{1}_L$ denotes an L-dimensional vector of ones and the last inequality follows from the definition of Λ_p. By the uniform Donsker theorem of Sheehy and Wellner (1992), we can show that for all $l \in \mathbb{N}_L$,

$$r_N \sup_{x\in\mathcal{X}} |\hat{v}_l(x) - v_l(x)| = O_P(1),\ P\text{-uniformly}. \tag{9.4}$$

The result (9.1) now follows from (9.3), (9.4) and the assumption $c_{N,1} \to \infty$.

We now establish (9.2). Let $\hat{\mathbf{s}}_A(x)$ be an L-dimensional vector whose l-th entry is $r_N\hat{v}_l(x)$ if $l \in A$ and $r_N\{\hat{v}_l(x) - v_l(x)\}$ if $l \in \mathbb{N}_L \setminus A$. We have

$$\begin{aligned}\int_{B_N(c_N)} \Lambda_{A,p}(\hat{\mathbf{u}}(x))dx &\leq \int_{B_N(c_N)} \Lambda_p(\hat{\mathbf{u}}(x))dx \\ &\leq \int_{B_N(c_N)} \Lambda_p(\hat{\mathbf{s}}_A(x) - c_{N,1}\mathbf{1}_{-L})dx,\end{aligned} \tag{9.5}$$

where $\mathbf{1}_{-L}$ denotes the L-dimensional vector whose l-th entry is zero if $l \in A$ and one if $l \in \mathbb{N}_L \setminus A$, the first and second inequalities holds by the definition of Λ_p and $B_{N,A}(c_N)$, respectively. Also, using (9.4) and the assumption $c_{N,1} \to \infty$, we have

$$\inf_{P \in \mathcal{P}_0} P \left\{ \int_{B_N(c_N)} \Lambda_p(\hat{\mathbf{s}}_A(x) - c_{N,1}\mathbf{1}_{-L})dx = \int_{B_N(c_N)} \Lambda_{A,p}(\hat{\mathbf{s}}_A(x))dx \right\} \to 1. \tag{9.6}$$

The result (9.2) now follows from (9.5) and (9.6) because

$$\int_{B_N(c_N)} \Lambda_{A,p}(\hat{\mathbf{s}}_A(x))dx = \int_{B_N(c_N)} \Lambda_{A,p}(\hat{\mathbf{u}}(x))dx.$$

□

Proof of Theorem 4: Let

$$\bar{T_N}(c_N) := \int_{B_N(c_N)} \Lambda_{A,p}(\hat{\mathbf{s}}(x))dx.$$

Under the null hypothesis, we have $v_l(x) \leq 0$ for all $l \in \mathbb{N}_L$. Therefore, Lemma 3 implies that

$$\inf_{P \in \mathcal{P}_0} P \left\{ T_N \leq \bar{T_N}(c_N) \right\} \to 1. \tag{9.7}$$

Let $\bar{c}^*_{N,\alpha}$ denote the $(1-\alpha)$ quantile of the bootstrap distribution of

$$\bar{T}^*_N(c_N) := \int_{B_N(c_N)} \Lambda_{A,p}(\hat{\mathbf{s}}^*(x))dx.$$

Under Assumptions 1 - 5, it can be shown that

$$\inf_{P \in \mathcal{P}} P \left\{ B_{N,A}(c_{N,1}, c_{N,2}) \subset \hat{B}_A(\hat{c}_N) \subset B_{N,A}(c_{N,2}, c_{N,2}) \right\} \to 1 \tag{9.8}$$

following the proof of Theorem 2, Claim 1 in Linton, Song and Whang (2010). By (9.8) and Assumption 5, with probability approaching one, we have

$$\bar{T}^*_N(c_N) \leq \sum_{A \in \mathcal{N}_L} \int_{\hat{B}_{N,A}(\hat{c}_N)} \Lambda_{A,p}(\hat{\mathbf{s}}^*(x))dx$$

and hence

$$\inf_{P \in \mathcal{P}} P \left\{ c^*_{N,\alpha} \geq \bar{c}^*_{N,\alpha} \right\} \to 1. \tag{9.9}$$

There exists a sequence of probabilities $\{P_N\}_{N_1,N_2 \geq 1} \subset \mathcal{P}_0$ such that

$$\begin{aligned} \limsup_{N_1,N_2 \to \infty} \sup_{P \in \mathcal{P}_0} P \left\{ T_N > c^*_{N,\alpha,\eta} \right\} &= \limsup_{N_1,N_2 \to \infty} P_N \left\{ T_N > c^*_{N,\alpha,\eta} \right\} \\ &= \lim_{N_1,N_2 \to \infty} P_{w_N} \left\{ T_{w_N} > c^*_{w_N,\alpha,\eta} \right\}, \end{aligned} \tag{9.10}$$

where $\{w_N := (w_{N_1}, w_{N_2})\}_{N_1,N_2\geq 1} \subset \{(N_1,N_2)\}_{N_1,N_2\geq 1}$ is a subsequence and T_{w_N} and $c^*_{w_N,\alpha,\eta}$ are the same as T_N and $c^*_{N,\alpha,\eta}$, respectively, except that the sample size N is replaced by w_N.

Let $\sigma_N(c_N) := \mathbf{Var}_P\left(\bar{T}_N(c_N)\right)$ denote the variance of $\bar{T}_N(c_N)$. By Assumption 1(c), $\{\sigma_N(c_N)\}_{N_1,N_2\geq 1}$ is a bounded sequence uniformly in $P \in \mathcal{P}$. Therefore, there exists a further subsequence $\{u_N := (u_{N_1}, u_{N_2})\}_{N_1,N_2\geq 1} \subset \{w_N\}_{N_1,N_2\geq 1}$ such that $\sigma_{u_N}(c_{u_N})$ converges. We will show below that

$$\limsup_{N_1,N_2\to\infty} P_{u_N}\left\{T_{u_N} > c^*_{u_N,\alpha,\eta}\right\} \leq \alpha. \tag{9.11}$$

Since $P_{w_N}\left\{T_{w_N} > c^*_{w_N,\alpha,\eta}\right\}$ converges along $\{w_N\}$, it also converges along the subsequence $\{u_N\}$. Therefore, the result of Theorem 4(i) holds because the limsup in (9.11) is equal to the limit in (9.10).

We now establish (9.11). Consider first the case $\lim_{N_1,N_2\to\infty}\sigma_{u_N}(c_{u_N}) > 0$. We have

$$\begin{aligned} P_{u_N}\left\{T_{u_N} > c^*_{u_N,\alpha,\eta}\right\} &\leq P_{u_N}\left\{T_{u_N} > \bar{c}^*_{u_N,\alpha}\right\} + o(1) \\ &\leq P_{u_N}\left\{\bar{T}_{u_N} > \bar{c}^*_{u_N,\alpha}\right\} + o(1) \\ &\leq \alpha + o(1), \end{aligned}$$

where the first inequality uses the fact that $c^*_{N,\alpha,\eta} \geq c^*_{N,\alpha} \geq \bar{c}^*_{N,\alpha}$ with probability approaching one by (9.9), the second inequality follows from (9.7), and the last inequality holds by the bootstrap consistency result in Lemma 9 and the uniform continuous mapping theorem (Linton, Song and Whang (2010, Lemma A1)).

We next consider the other case: $\lim_{N_1,N_2\to\infty}\sigma_{u_N}(c_{u_N}) = 0$. In this case, we have

$$\begin{aligned} P_{u_N}\left\{T_{u_N} > c^*_{u_N,\alpha,\eta}\right\} &\leq P_{u_N}\left\{\bar{T}_{u_N} > c^*_{u_N,\alpha,\eta}\right\} + o(1) \\ &\leq P_{u_N}\left\{\bar{T}_{u_N} > \eta\right\} + o(1) \\ &= o(1), \end{aligned}$$

where the first inequality follows from (9.7), the second inequality holds by the definition (4.8), and the last convergence to zero follows from the condition $\lim_{N_1,N_2\to\infty}\sigma_{u_N}(c_{u_N}) = 0$ and the fact $\eta > 0$. This completes the proof of Theorem 4. □

Proof of Theorem 5: By convexity of the map Λ_p, we have

$$T_N = \int_{\mathcal{X}} \Lambda_p(\hat{\mathbf{s}}(x) + \mathbf{u}(x))dx$$
$$\geq \frac{1}{2^{p-1}} \int_{\mathcal{X}} \Lambda_p(\mathbf{u}(x))dx - \int_{\mathcal{X}} \Lambda_p(-\hat{\mathbf{s}}(x))dx. \quad (9.12)$$

By (9.4), the last term of (9.12) is $O_P(1)$. Since $r_N \to \infty$ and $\int_{\mathcal{X}} \Lambda_p(v_1(x), \ldots, v_L(x))\, dx > 0$, we have for any constant $M_1 > 0$,

$$P\left\{\frac{1}{2^{p-1}} \int_{\mathcal{X}} \Lambda_p(\mathbf{u}(x))dx > M_1\right\} \to 1.$$

Therefore, this implies that for any constant $M_2 > 0$,

$$P\left\{T_N > M_2\right\} \to 1.$$

The result of Theorem 5 now holds because $c^*_{N,\alpha,\eta} = O_P(1)$ by Lemma 9. □

References

[1] Anderson, G. J. (1996). Nonparametric Tests of Stochastic Dominance in Income Distributions, *Econometrica*, 64, 1183-1193.

[2] Banerjee, A. and Duflo, E. (2014). Testing Credit Constraints, *Review of Economics Studies*, 81, 572-607.

[3] Banerjee, A. and Newman, A. (1993). Occupational Choice and the Process of Development, *Journal of Political Economy*, 101 274-298.

[4] Barrett, G. F. and Donald, S. G. (2003). Consistent Tests for Stochastic Dominance, *Econometrica*, 71, 71-104.

[5] Cheng, Q. and Z. Fang (2019). Inference on Functionals under First Order Degeneracy, *Journal of Econometrics* 201, 459-481.

[6] Cohen, J.D., Ericson, K.M., Laibson, D., and White , J.M. (2020). Measuring Time Preferences, *Journal of Economic Literature*, 58(2), 299-347.

[7] Davidson, R. and Duclos, J.-Y. (2000). Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica*, 68, 1435-1464.

[8] Dietz, S. and Matei, N. A. (2016). Spaces for Agreement: A Theory of Time-Stochastic Dominance and an Application to Climate Change, *Journal of the Association of Environmental and Resource Economists* 3, 85-130.

[9] Ekern, S. (1981). Time Dominance Efficiency Analysis, *Journal of Finance*, 36(5), 1023–33.

[10] Fang, Z. and Santos, A. (2018). Inference on Directionally Differentiable Functions, *Review of Economic Studies*, 86, 377-412.

[11] Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time Discounting and Time Preference: A Critical Review, *Journal of Economic Literature*, 40(2), 351-401.

[12] Galor, O. and Zeira, J. (1993). Income Distribution and Macroeconomics, *Review of Economics Studies*, 60, 35-52.

[13] Holland, S. P., Mansur, E. T., Muller, N. Z., and Yates, A. J. (2016). Are there Environmental Benefits from Driving Electric Vehicles? The Importance of Local Factors, *American Economic Review*, 106(12), 3700-3729.

[14] Hong, H. and Li, J. (2018). The Numerical Delta Method, *Journal of Econometrics* 206, 379-394.

[15] Kaboski, J. and Townsend, R. (2011). A Structural Evaluation of a Large-scale Quasi-Experimental Microfinance Initiative, *Econometrica*, 79(5), 1357-1406.

[16] Kaboski, J. and Townsend, R. (2012). The Impact of Credit on Village Economies, *American Economic Journal: Applied Economics*, 4(2), 98-133.

[17] Kaur, A., Prakasa Rao, B. L. S. and Singh, H. (1994). Testing for Second-Order Stochastic Dominance of Two Distributions, *Econometric Theory*, 10, 849-866.

[18] Klecan, L., McFadden, R., and McFadden, D. (1991). A Robust Test for Stochastic Dominance. Working Paper, Department of Economics, MIT.

[19] Lee, S., Song, K. and Whang, Y. -J. (2018). Testing for a General Class of Functional Inequalities, *Econometric Theory*, 34, 1018-1064.

[20] Levy, H. (1973). Stochastic Dominance, Efficiency Criteria, and Efficient Portfolios: The Multi-Period Case, *American Economic Review*, 63(5), 986–94.

[21] Levy, H. (1998). *Stochastic Dominance: Investment Decision Making under Uncertainty*, Kluwer Academic Publishers

[22] Levy, H. and Paroush. J. (1974). Multi-Period Stochastic Dominance, *Management Science*, 21(4), 428–35

[23] Linton, O., Maasoumi, E., and Whang, Y. -J. (2005). Consistent Testing for Stochastic Dominance under General Sampling Schemes, *Review of Economic Studies*, 72(3), 735-765.

[24] Linton, O., Song, K. and Whang, Y. -J. (2010). An Improved Bootstrap Test of Stochastic Dominance, *Journal of Econometrics*, 154, 186-202.

[25] McFadden, D. (1989). Testing for Stochastic Dominance, *Studies in the Economics of Uncertainty*, pp.113-134. Springer, New York, NY.

[26] Mehra, R., and Prescott, E.C. (1985). The Equity Premium: A Puzzle. *Journal of Monetary Economics*, 15(2), 145-161.

[27] Nordhaus, W. (2007). A Review of the Stern Review on the Economics of Climate Change, *Journal of Economic Literature*, 45(3), 686-702.

[28] Shapiro, A. (1990). On Concepts of Directional Differentiability, *Journal of Optimization Theory and Applications*, 66, 477-487.

[29] Sheehy, A. and Wellner, J. A. (1992). Uniform Donkser Classes of Functions, *Annals of Probability*, 20, 1983-2030.

[30] Stern, N. H. (2007). *The Economics of Climate Change: The Stern Review*, Cambridge University Press.

[31] Weitzman, M. L. (2007). A Review of the Stern Review on the Economics of Climate Change, *Journal of Economic Literature*, 45(3), 703-724.

[32] Whang, Y. -J. (2019). *Econometric Analysis of Stochastic Dominance: Concepts, Methods, Tools, and Applications*, Cambridge University Press.

Department of Economics, Seoul National University
Email address: kh.lee@snu.ac.kr

Department of Economics, University of Cambridge
Email address: obl20@cam.ac.uk

Department of Economics, Seoul National University
Email address: whang@snu.ac.kr