Statistica Sinica Preprint No: SS-2025-0398	
Title	Bootstrap Consistency for Empirical Likelihood in Density Ratio Models
Manuscript ID	SS-2025-0398
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0398
Complete List of Authors	Weiwei Zhuang, Weiqi Yang and Jiahua Chen
Corresponding Authors	Jiahua Chen
E-mails	jhchen@stat.ubc.ca
Notice: Accepted author version.	

Statistica Sinica

BOOTSTRAP CONSISTENCY FOR EMPIRICAL LIKELIHOOD IN DENSITY RATIO MODELS

Weiwei Zhuang[1], Weiqi Yang[1] and Jiahua Chen[2*]

[1] *School of Management, University of Science and Technology of China*

[2] *Department of Statistics, The University of British Columbia*

Abstract: We establish the validity of bootstrap methods for empirical likelihood (EL) inference under the density ratio model (DRM). In particular, we prove that the bootstrap maximum EL estimators share the same limiting distribution as their population counterparts, both at the parameter level and for distribution functionals. Our results extend existing pointwise convergence theory to weak convergence of processes, which in turn justifies bootstrap inference for quantiles and dominance indices within the DRM framework. These theoretical guarantees close an important gap in the literature, providing rigorous foundations for resampling-based confidence intervals and hypothesis tests. Simulation studies further demonstrate the accuracy and practical value of the proposed approach.

Key words and phrases: Bootstrap; Density ratio model; Empirical likelihood; Semiparametric inference; Quantile processes.

Mathematics Subject Classifications (2020): 62G09, 62E20, 62G10, 62G30.

*Corresponding author: Jiahua Chen Email: jhchen@stat.ubc.ca ORCID: 0000-0001-8064-4444

1. Introduction

The empirical likelihood (EL) provides a powerful framework for statistical inference without restrictive parametric assumptions. Since its inception (Owen, 1988), EL has been applied in linear models (Owen, 1991), estimating equations (Qin and Lawless, 1994), and sample surveys (Chen and Qin, 1993; Wu and Sitter, 2001). Reviews in economics (Parente and Smith, 2014), survey methodology (Wu and Lu, 2016), and the comprehensive account in Qin (2017) attest to its broad and continuing impact.

A particularly important setting for EL inference is the *density ratio model* (DRM), introduced by Anderson (1979) and further developed by Qin and Zhang (1997). The DRM has been widely studied (Qin, 1998; Fokianos et al., 2001; Chen and Liu, 2013; Cai et al., 2017; Zhang and Chen, 2022) and applied to time series (Kedem et al., 2008), mixture models (Tan, 2009; Li et al., 2017), regression (Huang and Rathouz, 2012), quality monitoring (Hu, 2018), and dominance indices (Zhuang et al., 2019). Compared with traditional approaches, EL methods under DRM are statistically efficient by combining information across multiple samples.

Maximum EL estimators under DRM are asymptotically normal with well-defined variance structures. Yet their variances depend on unknown features and are analytically complex, limiting practical use. This difficulty has motivated widespread application of the bootstrap. Since its inception by Efron (1979), the bootstrap method has evolved into a cornerstone of modern statistical inference, with comprehensive theoretical foundations detailed in classic monographs such as Efron and Tibshirani (1993), Shao and Tu (1995), and Davison and Hinkley (1997). In recent decades, bootstrap theory has been significantly advanced to address complex statistical challenges. A pivotal development is the unified framework established by Cheng and Huang (2010), which proved bootstrap consistency for general semiparametric M-estimation. Subsequent research has further extended inferential validity to high-dimensional and functional settings Horowitz and Krishnamurthy (2018); Chernozhuokov et al. (2022), as well as to data with complex dependence structures such as Markov chains Soukarieh and Bouzebda (2023). In this spirit, we extend bootstrap theory to DRM and provide rigorous justification tailored to its structural features.

This paper develops a rigorous theoretical framework for bootstrap inference under the density ratio model. We show that bootstrap estimators of both model parameters and distributional functionals share the same limiting distributions as their population counterparts. Our results strengthen pointwise convergence to weak convergence in function space, thereby supporting bootstrap inference for quantiles and stochastic dominance indices. Together, these findings provide a solid theoretical foundation for bootstrap-based confidence intervals and hypothesis testing within the DRM framework.

The remainder of the paper is organized as follows. Section 2 reviews the DRM framework, and Section 3 establishes bootstrap consistency results. Section 4 develops applications to quantiles and dominance indices. Sections 5 and 6 present simulation results and practical illustrations, respectively. Section 7 concludes. To save space, all proofs in Sections 3 and 4, as well as additional details for the real data example, are provided in the supplementary material.

2. Preliminary

Suppose $\{x_{kj}, j = 1, \ldots, n_k\}$ are independent and identically distributed (i.i.d.) observations from F_k, $k = 0, 1, \ldots, m$. Let $n = \sum_{k=0}^{m} n_k$ be the total sample size and $\rho_k = n_k/n$, $k = 0, 1, \ldots, m$. We assume these distributions are related through the density ratio model (DRM)

$$dF_k(x) = \exp\{\boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x)\}\, dF_0(x), \quad k = 1, \ldots, m. \tag{2.1}$$

Here $\mathbf{q}(\cdot)$ is a known vector-valued basis function and $\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_m^{\mathrm{T}})$ is an unknown vector of parameters. As customary under the DRM, $\boldsymbol{\theta}_0$ represents the tilt from the baseline distribution F_0 to itself. Since there is no tilt in this case, we fix $\boldsymbol{\theta}_0 = \mathbf{0}$ for identifiability and notational symmetry. Consequently, the corresponding estimator satisfies $\widehat{\boldsymbol{\theta}}_0 = \mathbf{0}$ by construction. Under the DRM, all distributions share the same support, a natural assumption in relevant applications.

Qin and Zhang (1997) and Qin (1998) appear to be the first to introduce EL to data analysis under the DRM. Let $p_{kj} = dF_0(x_{kj})$. Then the empirical likelihood under DRM is

$$L_n(F_0, \ldots, F_m) = \prod_{k,j} dF_k(x_{kj}) = \prod_{k,j} p_{kj} \times \exp\left\{\sum_{k,j} \boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x_{kj})\right\},$$

where the products and sums range over $k = 0, \ldots, m$ and $j = 1, \ldots, n_k$. The function L_n depends on $\boldsymbol{\theta}$ and F_0, and its logarithm is

$$l_n(\boldsymbol{\theta}, F_0) = \sum_{k,j} \log(p_{kj}) + \sum_{k,j} \boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x_{kj}).$$

Inference on $\boldsymbol{\theta}$ is typically derived by profiling the EL with respect to F_0. Following Chen and Liu (2013), we first maximize the EL with respect to p_{kj} subject to the normalization restriction $\sum_{k,j} p_{kj} = 1$ and the m model constraints

$$\sum_{k,j} p_{kj} \exp\{\boldsymbol{\theta}_r^{\mathrm{T}} \mathbf{q}(x_{kj})\} = 1, \quad r = 1, \ldots, m.$$

By the method of Lagrange multipliers, the implied probabilities are given by

$$p_{kj} = n^{-1}\left\{1 + \sum_{s=1}^{m} \nu_s \big[\exp(\boldsymbol{\theta}_s^{\mathrm{T}} \mathbf{q}(x_{kj})) - 1\big]\right\}^{-1},$$

where the vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)^{\mathrm{T}}$ consists of the Lagrange multipliers determined by the aforementioned constraints. Substituting these fitted values of p_{kj} back into the original log-EL function, we obtain the profile log-EL for $\boldsymbol{\theta}$:

$$\tilde{l}_n(\boldsymbol{\theta}) = -\sum_{k,j} \log\left\{1 + \sum_{s=1}^{m} \nu_s \big[\exp(\boldsymbol{\theta}_s^{\mathrm{T}} \mathbf{q}(x_{kj})) - 1\big]\right\} + \sum_{k,j} \boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x_{kj}).$$

It is straightforward to show that $\tilde{l}_n$ achieves the same maximum value, and at the same maximizer, as the simpler dual log-EL:

$$l_n(\boldsymbol{\theta}) = -\sum_{k,j} \log \left[\sum_{s=0}^{m} \rho_s \exp\{\boldsymbol{\theta}_s^{\mathrm{T}} \mathbf{q}(x_{kj})\} \right] + \sum_{k,j} \boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x_{kj}), \tag{2.2}$$

where $\rho_s = n_s/n$. The dual log-EL enjoys many of the same analytical properties as an ordinary parametric likelihood. Moreover, it is concave in $\boldsymbol{\theta}$, which ensures the uniqueness of its maximizer under the regularity conditions stated below. It often leads to asymptotically normal maximum EL estimators and likelihood ratio statistics with chi-squared limits, providing a convenient foundation for inference under the DRM.

The maximum empirical likelihood estimator (MELE) of $\boldsymbol{\theta}$ is the maximizer of (2.2), denoted $\widehat{\boldsymbol{\theta}}$. Given $\widehat{\boldsymbol{\theta}}$, the fitted values of p_{kj} are

$$\widehat{p}_{kj} = \{n \cdot h(x_{kj}; \widehat{\boldsymbol{\theta}})\}^{-1}, \quad h(x; \boldsymbol{\theta}) = \sum_{k=0}^{m} \rho_k \exp(\boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x)).$$

Consequently, the MELE of $F_r(x)$ is

$$\begin{aligned} \widehat{F}_r(x) &= \sum_{k,j} \widehat{p}_{kj} \exp(\widehat{\boldsymbol{\theta}}_r^{\mathrm{T}} \mathbf{q}(x_{kj})) \mathbf{I}(x_{kj} \le x) \\ &= n_r^{-1} \sum_{k,j} h_r(x_{kj}; \widehat{\boldsymbol{\theta}}) \mathbf{I}(x_{kj} \le x), \end{aligned}$$

where $\mathbf{I}(A)$ is the indicator of event A, $\widehat{\boldsymbol{\theta}}_0 = \mathbf{0}$, and

$$h_r(x; \boldsymbol{\theta}) = \rho_r \exp(\boldsymbol{\theta}_r^{\mathrm{T}} \mathbf{q}(x))/h(x; \boldsymbol{\theta}).$$

The MELE $\widehat{\boldsymbol{\theta}}$ is asymptotically normal under the following assumptions.

Assumption 1.

(i) The total sample size $n = \sum_{k=0}^{m} n_k \to \infty$, with proportions $\rho_k = n_k/n$ converging to constants for $k = 0, \ldots, m$.

(ii) $F_0, \ldots, F_m$ satisfy the DRM (2.1) with true parameter $\boldsymbol{\theta}^{\dagger}$, and $\int h(x; \boldsymbol{\theta})\, dF_0(x) < \infty$ for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}^{\dagger}$. The components of $\mathbf{q}(x)$ are algebraically independent, with the first component equal to one.

(iii) Each F_k has a bounded density f_k and finite second moment.

In words, the sample sizes from different populations are comparable, the basis functions $\mathbf{q}(x)$ contain no redundancy, and the relevant moments are finite. Let $\mathbf{W}$ and $\mathbf{S}$ be $md \times md$ block matrices, with each block of size $d \times d$, defined by

$$\mathbf{W}_{rs} = \int \mathbf{q}(x)\mathbf{q}(x)^{\mathrm{T}}\{h_r(x; \boldsymbol{\theta}^{\dagger})\delta_{rs} - h_r(x; \boldsymbol{\theta}^{\dagger})h_s(x; \boldsymbol{\theta}^{\dagger})\}\, d\bar{F}(x), \quad d\bar{F}(x) = h(x; \boldsymbol{\theta}^{\dagger})dF_0(x),$$

and

$$\mathbf{S}_{rs} = (\rho_r^{-1}\delta_{rs} + \rho_0^{-1})\,\mathrm{diag}\{1, 0, \ldots, 0\}, \quad 1 \le r, s \le m,$$

where $\delta_{rs} = 1$ if $r = s$ and 0 otherwise. It can be shown that $\mathbf{W}^{-1} - \mathbf{S}$ is positive definition under Assumption 1.

Lemma 1. *Under Assumption 1, the MELE is asymptotically normal: as $n \to \infty$,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\dagger}) \rightsquigarrow N(\mathbf{0}, \mathbf{W}^{-1} - \mathbf{S})$$

In light of this asymptotic normality, one could in principle construct confidence regions for $\boldsymbol{\theta}$ or test hypotheses such as equality of distributions. However, such procedures require explicit evaluation and consistent estimation of the scaling matrix $\mathbf{W}^{-1} - \mathbf{S}$, whose analytic form is intricate and depends on unknown population quantities. An appealing alternative is to employ bootstrap resampling. Let $\widehat{\boldsymbol{\theta}}_b^*,\ b = 1, 2, \ldots, B$, denote bootstrap replicates obtained by resampling within each group. If the bootstrap is valid, then the conditional distribution (given the data) of

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$$

approximates the distribution of

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger),$$

thereby providing inference without the need to explicitly estimate $\mathbf{W}^{-1} - \mathbf{S}$.

Although bootstrap methods perform well in simulations, rigorous justification is essential. This paper takes up that task and contributes in three directions:

1. **Bootstrap consistency of parameter estimation**: we show that $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$ given the data and $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger)$ have the same limiting distribution.

2. **Bootstrap consistency of distribution estimation**: letting $\widehat{F}_r^*(x)$ denote the bootstrap MELE of $F_r(x)$, we show that $\sqrt{n}(\widehat{F}_r^*(x) - \widehat{F}_r(x))$ given the data and $\sqrt{n}(\widehat{F}_r(x) - F_r(x))$ share the same limiting process.

3. **Validation of existing bootstrap procedures**: our general results confirm the theoretical validity of many bootstrap inference methods for the DRM proposed in the literature.

We present these results in the next section.

3. Main Results

We begin by introducing the function spaces and notation used throughout. Let $l^\infty(\mathbb{R})$ denote the space of bounded, real-valued functions on $\mathbb{R}$, equipped with the supremum norm $\|g\|_\infty = \sup_{x\in\mathbb{R}} |g(x)|$. Let $L_1[0,1]$ denote the space of Lebesgue integrable functions on $[0,1]$, equipped with the L_1 seminorm $\|g\|_1 = \int_0^1 |g(x)|\, dx$. The expectation and variance of $g(X)$ when X has distribution F_k is denoted as $\mathrm{E}_k[g(X)]$ and $\mathrm{Var}_k[g(X)]$. We write $\rightsquigarrow$ for weak convergence.

3.1 Bootstrap Estimators of Parameters

We consider the following resampling scheme. For each $k = 0, \ldots, m$, draw a bootstrap sample $\mathcal{X}_k^* = \{x_{kj}^*, j = 1, \ldots, n_k\}$ with replacement from $\mathcal{X}_k = \{x_{kj}, j = 1, \ldots, n_k\}$. Let $p_{kj}^* = dF_0(x_{kj}^*)$. The bootstrap log-EL is

$$l_n^*(\boldsymbol{\theta}, F_0) = \sum_{k,j} \log(p_{kj}^*) + \sum_{k,j} \boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x_{kj}^*),$$

and the corresponding bootstrap dual log-EL is

$$l_n^*(\boldsymbol{\theta}) = -\sum_{k,j} \log\left\{\sum_{s=0}^{m} \rho_s \exp\big(\boldsymbol{\theta}_s^{\mathrm{T}} \mathbf{q}(x_{kj}^*)\big)\right\} + \sum_{k,j} \boldsymbol{\theta}_k^{\mathrm{T}} \mathbf{q}(x_{kj}^*). \tag{3.3}$$

The bootstrap MELE $\widehat{\boldsymbol{\theta}}^*$ maximizes (3.3).

Next we show that $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger)$ have the same limiting distribution.

Theorem 1. *Under Assumption 1,*

$$\sup_{\boldsymbol{\theta}} \left| \mathrm{P}\Big(\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) \le \boldsymbol{\theta} \,\Big|\, \mathcal{X}_n\Big) - \mathrm{P}(\boldsymbol{Z} \le \boldsymbol{\theta}) \right| = \mathbf{o}_p(1), \tag{3.4}$$

where $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbf{W}^{-1} - \mathbf{S})$.

Proof. The proofs of this theorem and a preparation lemma are given in the supplementary material.

3.2 Bootstrap Estimators of Distribution Functions

Chen and Liu (2013) established the asymptotic distribution of $\sqrt{n}\,\big(\widehat{F}_r(x) - F_r(x)\big)$ for fixed x. We extend this pointwise convergence to weak convergence of processes.

Building on the parameter expansion, the estimator $\widehat{F}_r$ admits a linearization that separates an empirical term from a smooth plug-in term involving $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger$. The former yields an empirical-process component, while the latter is controlled by the asymptotic normality of $\widehat{\boldsymbol{\theta}}$. This decomposition leads from pointwise limits to weak convergence of the entire distribution-function process.

Theorem 2. *Under Assumption 1, for* $r = 0, \ldots, m$,

$$\sqrt{n}\left(\widehat{F}_r(\cdot) - F_r(\cdot)\right) \rightsquigarrow \mathcal{G}_r(\cdot) \quad in\ l^{\infty}(\mathbb{R}),$$

where $\mathcal{G}_r$ *is a mean-zero Gaussian process with covariance function*

$$\omega_r(x,y) = \sigma_r(x,y) - \rho_r^{-2}\Big\{a_r(x \wedge y) - \mathbf{B}_r^{\mathrm{T}}(x)\,\mathbf{W}^{-1}\,\mathbf{B}_r(y)\Big\}, \tag{3.5}$$

with $x \wedge y = \min\{x, y\}$ *and*

$$\sigma_r(x,y) = \rho_r^{-1}\big\{F_r(x \wedge y) - F_r(x)F_r(y)\big\}, \qquad a_r(x) = \int_{-\infty}^{x} \big\{h_r(t;\boldsymbol{\theta}^{\dagger}) - h_r^2(t;\boldsymbol{\theta}^{\dagger})\big\}\, d\bar{F}(t),$$

and where $\mathbf{B}_r(x)$ *is an md-vector with sth d-segment*

$$\mathbf{B}_{r,s}(x) = \int_{-\infty}^{x} \big\{\delta_{rs}\, h_r(t;\boldsymbol{\theta}^{\dagger}) - h_r(t;\boldsymbol{\theta}^{\dagger})h_s(t;\boldsymbol{\theta}^{\dagger})\big\}\, \mathbf{q}(t)\, d\bar{F}(t), \quad s = 1, \ldots, m.$$

Proof. The proof of this theorem proceeds in three steps: establishing asymptotic joint normality, working out the covariance function $\omega_r(x, y)$, and verifying tightness. Together, these results establish the claimed weak convergence. The details are given in the supplementary material.

Remark 1. The covariance formula (3.5) can be interpreted as follows. The leading term $\sigma_r(x, y)$ is the natural covariance of the empirical distribution function based on group-r observations. The adjustment term

$$\rho_r^{-2}\Big\{a_r(x \wedge y) - \mathbf{B}_r^{\mathrm{T}}(x)\,\mathbf{W}^{-1}\,\mathbf{B}_r(y)\Big\}$$

arises from the fact that $\widehat{F}_r$ depends on the estimated parameter $\widehat{\boldsymbol{\theta}}$. The component $a_r(\cdot)$ accounts for additional sampling variability introduced through h_r, while the quadratic form $\mathbf{B}_r^{\mathrm{T}}(x)\mathbf{W}^{-1}\mathbf{B}_r(y)$ reflects the variance reduction due to joint estimation of $\boldsymbol{\theta}$. Thus (3.5) captures the balance between empirical fluctuations of F_r and the stabilizing effect of parameter estimation.

The bootstrap estimator of F_r is

$$\widehat{F}_r^*(x) = n_r^{-1} \sum_{k,j} h_r(x_{kj}^*; \widehat{\boldsymbol{\theta}}^*)\, \mathbf{I}(x_{kj}^* \leq x). \tag{3.6}$$

The next theorem shows that the bootstrap process $\sqrt{n}\left(\widehat{F}_r^*(\cdot) - \widehat{F}_r(\cdot)\right)$ shares the same weak limit as in Theorem 2, thereby justifying the use of resampling-based confidence regions.

Theorem 3. *Under the conditions of Theorem 1, for $r = 0, \ldots, m$, conditionally on $\mathcal{X}_n$,*

$$\sqrt{n}\left(\widehat{F}_r^*(\cdot) - \widehat{F}_r(\cdot)\right) \rightsquigarrow \mathcal{G}_r(\cdot) \quad \text{in } l^\infty(\mathbb{R}),$$

where $\mathcal{G}_r$ is as in Theorem 2.

Proof. The argument parallels that of Theorem 2. In supplementary material, we will highlight the main steps of the proof.

Remark 2. Theorem 3 establishes that the bootstrap reproduces the same Gaussian limit as the original process. This guarantees that confidence bands or regions for F_r constructed by resampling are asymptotically valid, providing a practical inference tool that requires no further analytic approximation beyond Theorem 2.

4. Applications

We apply the preceding theory to several inference problems under the density ratio model (DRM), showing that the bootstrap procedures used in prior work are theoretically valid and consistent with our results.

4.1 Quantile Functions

We define the quantile function associated with a cumulative distribution function F by $Q(p) = \inf x : F(x) \geq p$ for $p \in (0, 1)$. For each population r under the DRM, let $Q_r(p)$ denote the corresponding quantile function of F_r. We write $\widehat{Q}_r(p)$ and $\widehat{Q}^*_r(p)$ for the DRM estimator and its bootstrap analogue, respectively. While Chen and Liu (2013) obtained the limiting distribution of $\widehat{Q}_r(p)$ (via a Bahadur representation) at fixed p, our framework extends this to the entire quantile process. Building on Theorem 2, we establish weak convergence of $\sqrt{n}(\widehat{Q}_r - Q_r)$ and its bootstrap analog in $L_1[0, 1]$.

Context. Working in $L_1[0, 1]$ is both natural and useful: for unbounded supports, Q_r itself can be unbounded, yet most target statistics are integrals of Q_r over $p \in (0, 1)$. Weak convergence in $L_1[0, 1]$ is therefore sufficient (via the continuous mapping theorem) for inference on such functionals.

A direct application of the functional delta method from $l^\infty(\mathbb{R})$ to $L_1[0,1]$ is obstructed because the inverse map $\phi : F \mapsto Q$ is not Hadamard differentiable under the sup norm (Kaji, 2018). To circumvent this, we proceed in two steps. First, we strengthen Theorem 2 to obtain weak convergence of $\sqrt{n}(\widehat{F}_r - F_r)$ in the space

$$\mathbb{L} = \Big\{ f : \mathbb{R} \to \mathbb{R} \text{ bounded and integrable} \Big\}, \qquad \|f\|_{\mathbb{L}} := \|f\|_\infty + \int_{\mathbb{R}} |f(x)|\, dx.$$

Second, we invoke Kaji (2018), Theorem 1.3, which shows that ϕ is Hadamard differentiable from $\mathbb{L}$ to $L_1[0,1]$, thereby enabling the functional delta method. The bootstrap result follows analogously.

Theorem 4. *Under the conditions of Theorem 1, for $r = 0, \ldots, m$ and $p \in (0,1)$,*

$$\sqrt{n}\left(\widehat{Q}_r(p) - Q_r(p)\right) \rightsquigarrow -Q_r'(p)\, \mathcal{G}_r\big(Q_r(p)\big) \quad \text{in } L_1[0,1], \tag{4.7}$$

where $Q_r'(p)$ is the derivative of $Q_r(p)$, and $\mathcal{G}_r$ is from Theorem 2. Furthermore, conditionally on $\mathcal{X}_n$,

$$\sqrt{n}\left(\widehat{Q}^*_r(p) - \widehat{Q}_r(p)\right) \rightsquigarrow -Q_r'(p)\, \mathcal{G}_r\big(Q_r(p)\big) \quad \text{in } L_1[0,1].$$

Proof. The proof of this result follows the techniques in Kaji (2018) and Kosorok (2008). The specific details are in the supplementary material.

Remark 3. Theorem 4 extends classical quantile asymptotics to the DRM setting at the process level. The appearance of $-Q_r'(p)\,\mathcal{G}_r(Q_r(p))$ in (4.7) mirrors the familiar Bahadur representation, but now the result holds uniformly in p and in the L_1 sense. The bootstrap statement ensures that resampling-based inference for quantiles inherits the same validity as for distribution functions, making procedures such as percentile bootstrap confidence bands theoretically justified.

4.2 Dominance Index

Let F_0 and F_1 be two DRM distributions with densities f_0, f_1 and quantile functions Q_0, Q_1. The dominance index of F_0 over F_1 is

$$\gamma^{\dagger} := \gamma(F_0, F_1) = \mu\{\, p \in (0,1) : Q_0(p) > Q_1(p)\,\},$$

where μ is Lebesgue measure. As it aggregates where F_0 exceeds F_1 across quantile levels, $\gamma^{\dagger}$ is robust to extremes and interpretable through stochastic-dominance intuition. When F_0 and F_1 cross finitely many times, Zhuang et al. (2019) consider the plug-in estimator

$$\widehat{\gamma} = \gamma(\widehat{F}_0, \widehat{F}_1) = \mu\{\, p \in (0,1) : \widehat{Q}_0(p) > \widehat{Q}_1(p)\,\}.$$

Our objective is to show that the limiting distribution of $\widehat{\gamma}$ coincides with that of its bootstrap analogue. A central technical difficulty is that the dominance index functional $\gamma^{\dagger}$ is not continuous in the quantile functions Q_0 and Q_1. This lack of continuity prevents a direct application of the standard functional Delta method to transfer asymptotic results from the quantile processes. To overcome this obstacle, we follow the proof strategy of Zhuang et al. (2019), leveraging the established consistency and distributional equivalence of the bootstrap quantile estimator to justify the inference.

Lemma 2. *(Zhuang et al., 2019) Assume $F_0(x) = F_1(x)$ at finitely many x, with $f_0(x) \neq f_1(x)$ at each crossing. Let $[x_L, x_U]$ be the common support of F_0 and F_1, and suppose*

$$\lim_{x \to x_L} \frac{f_1(x)}{f_0(x)} \neq 1, \qquad \lim_{x \to x_U} \frac{f_1(x)}{f_0(x)} \neq 1. \tag{4.8}$$

Then $n^{1/2}\{\widehat{\gamma} - \gamma^{\dagger}\} \rightsquigarrow N(0, \sigma^2)$ for a variance σ^2 (available in closed form, though algebraically involved).

Remark (motivation). While Lemma 2 yields a CLT, σ^2 depends on unknown features of (F_0, F_1) and is cumbersome to estimate directly. Leveraging our process-level results for F_r and Q_r (Theorems 2–4), we instead validate a bootstrap that *automatically* reproduces the limit law of $\widehat{\gamma}$, avoiding explicit variance estimation.

Define the bootstrap estimator

$$\widehat{\gamma}^* = \gamma(\widehat{F}_0^*, \widehat{F}_1^*) = \mu\{\, p \in (0,1) : \widehat{Q}_0^*(p) > \widehat{Q}_1^*(p)\, \},$$

where $\widehat{F}_r^*, \widehat{Q}_r^*$ are the DRM bootstrap counterparts.

Theorem 5. *Under the conditions of Lemma 2,*

$$\sup_x \left| \mathrm{P}\big(\sqrt{n}(\widehat{\gamma}^* - \widehat{\gamma}) \le x \,\big|\, \mathcal{X}_n\big) - \mathrm{P}(Z_\gamma \le x) \right| = \mathbf{o}_p(1), \tag{4.9}$$

where $Z_\gamma \sim N(0, \sigma^2)$ and σ^2 is as in Lemma 2.

Proof. The proof is given in the supplementary material.

Remark 4. Theorem 5 confirms that the bootstrap reproduces the sampling law of $\widehat{\gamma}$ without explicit estimation of the intricate variance in Lemma 2. In practice, percentile or basic bootstrap intervals for $\gamma(F_0, F_1)$ are therefore asymptotically valid, and the approach aligns seamlessly with the process-level results established in Theorems 2–4.

5. Simulation

We assess the finite-sample performance of the bootstrap procedures developed in Theorems 2–4. Specifically, we examine the empirical coverage of percentile bootstrap confidence intervals (CIs) for (i) the DRM parameter vector $\boldsymbol{\theta}$, (ii) the distribution value $F_r(x)$ at selected x, and (iii) the quantile $Q_r(p) = F_r^{-1}(p)$ at selected p, targeting the nominal level $1 - \alpha = 0.95$.

Generic target and bootstrap CI. Let ξ be any *scalar functional* of $(F_0, \ldots, F_m)$–for example, a component θ_{rs} of $\boldsymbol{\theta}$, a value $F_r(x)$ at a fixed x, or a quantile $F_r^{-1}(p)$ at a fixed p. Denote its true value by $\xi^\dagger$, the DRM estimator by $\widehat{\xi}$, and the bth bootstrap replicate by $\widehat{\xi}^{*b}$. By the bootstrap consistency established earlier, the laws of $\widehat{\xi}-\xi^{\dagger}$ and $\widehat{\xi}^{*}-\widehat{\xi}$ share the same limit. If q_α denotes the empirical α–quantile of $\{\widehat{\xi}^{*b}-\widehat{\xi} : b=1,\ldots,B\}$, then

$$\mathrm{P}\big(\widehat{\xi}-\xi^{\dagger}\le q_\alpha\big)\approx\alpha.$$

This yields the standard *percentile* CI for ξ:

$$\mathrm{CI}_{1-\alpha}(\xi)=\Big[\,\widehat{\xi}-q_{1-\alpha/2}\,,\;\widehat{\xi}-q_{\alpha/2}\,\Big]. \tag{5.10}$$

Design, resampling and evaluation. In each scenario we generate i.i.d. observations $\{x_{rj}\}_{j=1}^{n_r}\sim F_r$ for groups $r=0,\ldots,m$, collected as $\mathcal{X}_n=\{x_{rj}\}$. For each Monte Carlo run, we compute $\widehat{\xi}$ for the following targets

- DRM parameters θ_{rs} with $r=1,\ldots,m$ and $s=1,\ldots,d$;
- distribution values $F_r(x)$ at $x=F_0^{-1}(p):p=0.1,\ldots,0.9$;
- quantiles $Q_r(p)=F_r^{-1}(p)$ at $p=0.1,\ldots,0.9$.

To approximate sampling variability, we perform $B=999$ bootstrap resamples by drawing with replacement within each group r (preserving $\{n_r\}$), recompute $\widehat{\xi}^{*b}$, and form the endpoints in (5.10). Empirical coverage rates are based on $N=2000$ Monte Carlo runs.

Scenarios and DRM bases. We consider two data–generating settings chosen to probe distinct shapes and tail behaviors.

1. *Gamma family* ($m = 4$)*:* $F_r \in \{\Gamma(5, 1.5), \Gamma(5, 1.4), \Gamma(6, 1.3), \Gamma(6, 1.2), \Gamma(7, 1.1)\}$ with sample sizes $(500, 450, 550, 650, 675)$; DRM basis $\mathbf{q}(x) = (1,\, x,\, \log x)^{\top}$ ($d = 3$).

2. *Normal family* ($m = 6$)*:* $F_r \in \{N(11, 1), N(11.5, 2), N(12, 3), N(12.5, 4), N(13, 5), N(13.5, 6), N(14, 7)\}$ with sample sizes $(300, 320, 340, 330, 350, 370, 400)$; DRM basis $\mathbf{q}(x) = (1,\, x,\, x^2)^{\top}$ ($d = 3$).

Here $N(\mu, \sigma^2)$ has mean μ and variance σ^2, and $\Gamma(\alpha, \beta)$ has shape α and scale β. Both families and their associated basis functions satisfy the conditions of the DRM.

Findings.

The simulation results are reported in Tables 16. Across both scenarios, the bootstrap confidence intervals (CIs) for the *model parameters* θ_{rs} and the *distribution values* $F_r(x)$ achieve empirical coverage close to the nominal 0.95 level. For the *quantiles* $Q_r(p)$, coverage remains near nominal, with mild undercoverage at extreme levels (e.g., $p \in 0.1, 0.9$ under the Normal setting), a well-known finite-sample phenomenon in quantile inference.

Tables 3–6 further compare the proposed bootstrap CIs with those based on the asymptotic theory of Chen and Liu (2013). The two approaches yield highly comparable coverage probabilities and average interval lengths. Although the bootstrap entails additional computational effort due to resampling, it relieves the practitioner from estimating the scaling matrix and other analytically intricate covariance components required by the asymptotic method. Such estimation involves substantial derivation and careful implementation. In this sense, the bootstrap exchanges modest computational cost for a meaningful reduction in analytical and coding burden, while maintaining competitive finite-sample performance.

Table 1: Empirical coverage of percentile bootstrap CIs for $\boldsymbol{\theta}$

Scenario I: Gamma ($d = 3$)

	F_1	F_2	F_3	F_4
θ_{r1}	.956	.945	.948	.942
θ_{r2}	.953	.948	.946	.938
θ_{r3}	.957	.944	.952	.941

Scenario II: Normal ($d = 3$)

	F_1	F_2	F_3	F_4	F_5	F_6
θ_{r1}	.959	.952	.945	.941	.943	.944
θ_{r2}	.957	.953	.943	.948	.950	.949
θ_{r3}	.955	.951	.942	.947	.954	.949

Table 2: Empirical coverage of percentile bootstrap CIs for $F_r(x)$

	Scenario I: Gamma ($d=3$)					Scenario II: Normal ($d=3$)						
	F_0	F_1	F_2	F_3	F_4	F_0	F_1	F_2	F_3	F_4	F_5	F_6
$x = Q_0(0.1)$	.938	.941	.916	.917	.944	.926	.935	.938	.943	.932	.933	.922
$x = Q_0(0.2)$	.939	.951	.933	.929	.915	.934	.938	.943	.945	.941	.936	.931
$x = Q_0(0.3)$	.949	.948	.932	.934	.927	.936	.945	.942	.948	.944	.943	.939
$x = Q_0(0.4)$	.946	.952	.935	.940	.927	.937	.944	.945	.945	.946	.945	.944
$x = Q_0(0.5)$	.950	.954	.940	.945	.936	.942	.945	.951	.951	.949	.952	.950
$x = Q_0(0.6)$	.942	.949	.938	.943	.953	.944	.946	.952	.945	.946	.953	.954
$x = Q_0(0.7)$	.941	.946	.944	.947	.948	.941	.949	.950	.948	.947	.954	.953
$x = Q_0(0.8)$	.946	.940	.947	.944	.951	.942	.946	.949	.948	.944	.955	.950
$x = Q_0(0.9)$	.943	.936	.951	.953	.947	.940	.942	.945	.946	.944	.958	.954

Table 3: Empirical coverage of CIs for $Q_r(p)$ (Scenario I)

	Bootstrap method					Method in Chen and Liu (2013)				
	F_0	F_1	F_2	F_3	F_4	F_0	F_1	F_2	F_3	F_4
$p = 0.1$	.945	.943	.951	.949	.945	.956	.954	.950	.954	.957
$p = 0.2$	.948	.941	.941	.938	.937	.952	.952	.952	.952	.955
$p = 0.3$	.935	.953	.934	.950	.937	.949	.953	.954	.958	.952
$p = 0.4$	.935	.937	.941	.949	.936	.950	.958	.957	.954	.953
$p = 0.5$	.936	.937	.938	.933	.944	.946	.956	.957	.950	.954
$p = 0.6$	.934	.939	.932	.938	.940	.938	.954	.951	.955	.955
$p = 0.7$	.953	.939	.935	.933	.952	.945	.950	.952	.954	.948
$p = 0.8$	.939	.942	.949	.951	.949	.940	.944	.952	.948	.946
$p = 0.9$	.939	.948	.941	.953	.946	.941	.944	.942	.944	.941

Table 4: Average length of CIs for $Q_r(p)$ (Scenario I)

	Bootstrap method					Method in Chen and Liu (2013)				
	F_0	F_1	F_2	F_3	F_4	F_0	F_1	F_2	F_3	F_4
$p = 0.1$	.239	.263	.289	.288	.345	.247	.271	.295	.295	.354
$p = 0.2$	.239	.265	.286	.286	.338	.249	.275	.297	.297	.351
$p = 0.3$	.246	.271	.292	.293	.345	.254	.281	.302	.303	.358
$p = 0.4$	.254	.283	.304	.305	.363	.262	.291	.313	.315	.374
$p = 0.5$	.269	.300	.322	.325	.389	.275	.306	.330	.334	.399
$p = 0.6$	.291	.325	.349	.355	.428	.295	.329	.356	.361	.435
$p = 0.7$	.326	.363	.392	.398	.481	.327	.366	.395	.401	.485
$p = 0.8$	.383	.430	.459	.465	.565	.381	.429	.458	.464	.564
$p = 0.9$	.501	.562	.594	.605	.745	.494	.555	.585	.595	.729

Table 5: Empirical coverage of CIs for $Q_r(p)$ (Scenario II)

	Bootstrap method							Method in Chen and Liu (2013)						
	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_0	F_1	F_2	F_3	F_4	F_5	F_6
$p = 0.1$	.946	.937	.949	.946	.948	.949	.937	.947	.952	.944	.951	.949	.956	.946
$p = 0.2$	.954	.939	.954	.950	.950	.948	.936	.953	.948	.955	.947	.958	.951	.945
$p = 0.3$	.959	.943	.953	.950	.950	.954	.943	.949	.954	.946	.952	.950	.957	.948
$p = 0.4$	.963	.947	.952	.952	.953	.954	.947	.956	.949	.953	.948	.952	.950	.955
$p = 0.5$	.961	.950	.954	.953	.950	.953	.948	.951	.947	.956	.954	.948	.952	.949
$p = 0.6$	.962	.951	.951	.957	.948	.951	.948	.955	.950	.947	.956	.953	.948	.952
$p = 0.7$	.956	.953	.948	.950	.952	.953	.955	.948	.956	.951	.949	.954	.955	.947
$p = 0.8$	.952	.947	.943	.950	.945	.952	.950	.952	.946	.955	.951	.947	.953	.950
$p = 0.9$	.946	.933	.940	.933	.934	.940	.941	.954	.949	.952	.946	.948	.951	.945

Table 6: Average length of CIs for $Q_r(p)$ (Scenario II)

	Bootstrap method							Method in Chen and Liu (2013)						
	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_0	F_1	F_2	F_3	F_4	F_5	F_6
$p = 0.1$	.321	.435	.514	.598	.650	.694	.728	.323	.439	.519	.603	.657	.703	.736
$p = 0.2$	.281	.381	.450	.523	.568	.606	.632	.277	.378	.446	.519	.564	.601	.626
$p = 0.3$	.261	.353	.417	.486	.529	.565	.591	.256	.346	.409	.476	.520	.555	.579
$p = 0.4$	.251	.338	.401	.468	.511	.547	.575	.244	.331	.390	.457	.498	.533	.562
$p = 0.5$	.247	.333	.396	.463	.506	.544	.575	.240	.324	.386	.451	.492	.531	.558
$p = 0.6$	.248	.337	.401	.469	.514	.553	.589	.242	.328	.390	.457	.502	.537	.574
$p = 0.7$	.256	.350	.416	.488	.534	.577	.618	.251	.342	.406	.479	.522	.566	.604
$p = 0.8$	.274	.375	.448	.524	.575	.622	.668	.272	.370	.443	.517	.567	.615	.662
$p = 0.9$	.311	.426	.508	.596	.656	.714	.771	.315	.430	.510	.603	.662	.726	.778

6. Real-data analysis

We illustrate resampling-based inference under the DRM using per capita income data from the 2020 wave of the China Family Panel Studies (CFPS). The CFPS is a nationally representative longitudinal survey designed to collect community-, family-, and individual-level information reflecting Chinas social, economic, demographic, educational, and health developments. The baseline survey covers 25 provinces, representing approximately 95% of the national population. Per capita income is constructed as total household gross (or net) income divided by household size. Household income aggregates wage income, operating income, property income, transfer income, and other sources, while household size is defined as the number of co-resident members.

We analyze six provinces: Henan ($n = 1160$), Hubei ($n = 163$), Hunan ($n = 299$), Fujian ($n = 166$), Anhui ($n = 223$), and Sichuan ($n = 510$). These provinces exhibit broadly comparable levels of economic development while spanning Central, East, and Southwest China, providing a meaningful setting to assess the practical performance of the proposed method. Each province is treated as a distinct population under the DRM framework.

A key advantage of employing the DRM in this context is its ability to accommodate substantial disparities in sample sizes across populations. As noted,

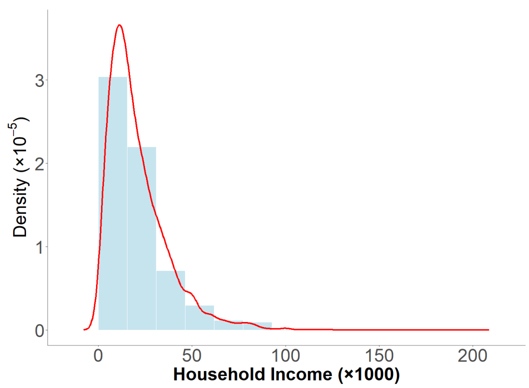
China Family Panel Studies (CFPS, 2020). See Institute of Social Science Survey (2015) the sample sizes range from $n = 163$ for Hubei to $n = 1160$ for Henan. A purely nonparametric analysis would produce estimates (e.g., poverty rates or medians) with markedly different levels of precision, thereby complicating meaningful cross-province comparisons. The DRM framework mitigates this issue by jointly estimating the distributions and effectively borrowing strength from larger samples to stabilize inference for smaller ones. By leveraging information from all six provinces under a common structural assumption, we obtain more comparable and statistically efficient inferences for each population.

We postulate the DRM with the basis

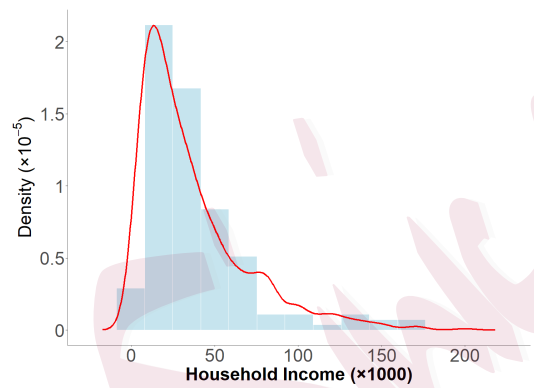
$$\mathbf{q}(x) = \left(1,\ x,\ x^2,\ \log x,\ \sqrt{x}\right)^{\mathrm{T}},$$

which is sufficiently flexible to capture Gaussian- and Gamma-like shapes commonly observed in income distributions. We fit the DRM to these six provinces and obtain fitted distribution $\widehat{F}_r(x)$, $r = 1, \ldots, 6$. For visualization, Figure 1 displays, for each province, a histogram of income (x-axis labeled in thousands of yuan) overlaid with the *DRM-implied* kernel density (convolution of $\widehat{F}_r$ with a Gaussian kernel). The DRM density estimates clearly match the histograms well, empirically supporting the DRM assumption. For readers interested in rigorous model validation, formal goodness-of-fit tests for the DRM are available in Qin and Zhang (1997); Bondell (2007).

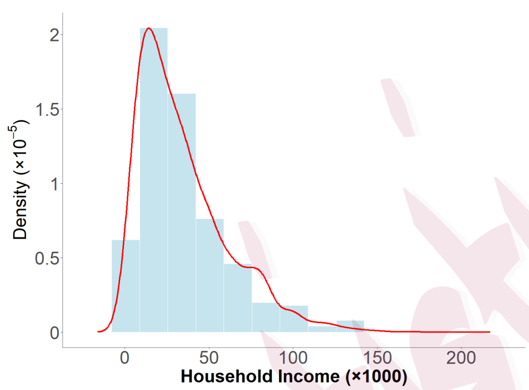
We now examine poverty rates, median incomes, and distributional dominance



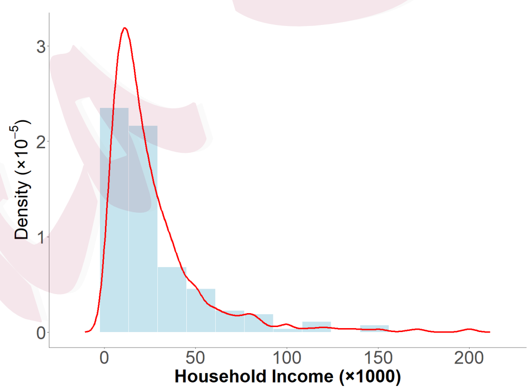
(a) Henan



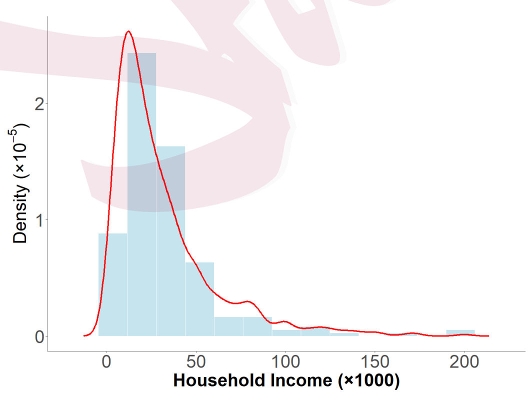
(b) Hubei



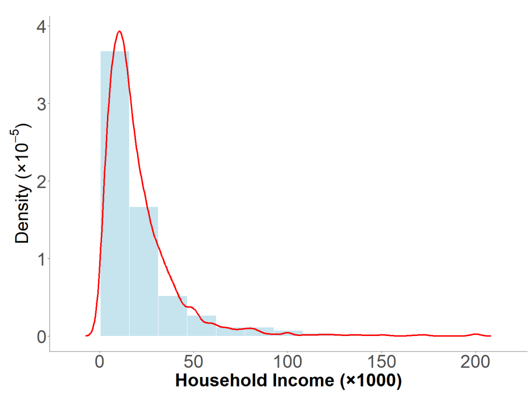
(c) Hunan



(d) Fujian



(e) Anhui



(f) Sichuan

Figure 1: Income histograms overlaid with DRM-implied density estimates.

across the six provinces, as these measures provide policy-relevant insights. Bootstrap methods are used to construct two-sided 95% confidence intervals for each quantity of interest. Specifically, we consider:

- **Poverty rate.** The poverty line is set at an annual income of 4,000 RMB in 2020. The corresponding poverty rate is therefore $F(4000)$.
- **Median income.** The median income is given by $Q(0.5)$.
- **Dominance index.** The dominance index of province r over province s is defined as

$$\gamma(F_r, F_s) \;=\; \mu\{\, p \in (0,1) : Q_r(p) > Q_s(p) \,\},$$

  where μ denotes the Lebesgue measure on $(0,1)$. A value exceeding $1/2$ indicates that more individuals in province r have experienced income improvement than those in province s.

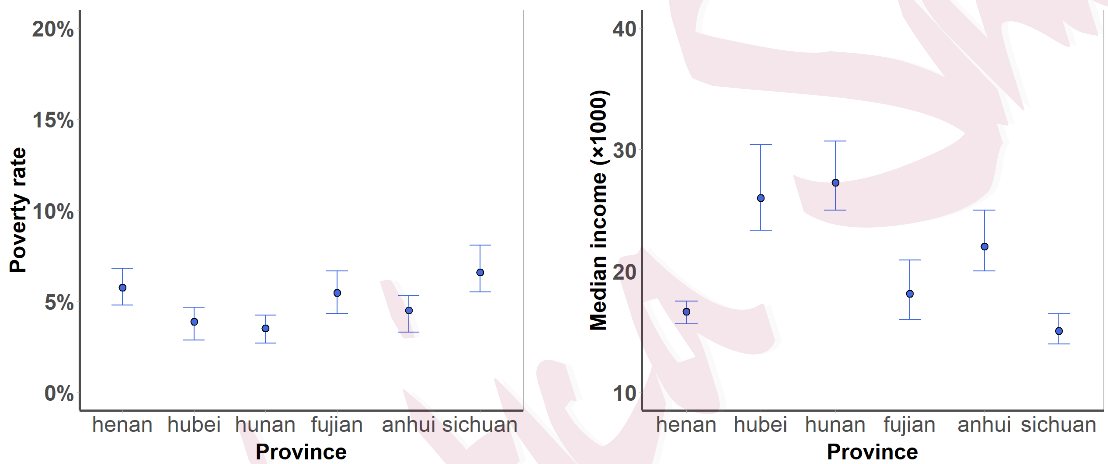
We construct percentile bootstrap 95% confidence intervals (CIs) as described in (5.10), using $B = 20001$ bootstrap replicates. This resampling procedure preserves the joint estimation structure imposed by the DRM and is theoretically justified by Theorems 3–4.

Figure 2 displays two forest-style panels:

(a) Poverty rates for the six provinces with 95% CIs.

(b) Median incomes with 95% CIs.

In each panel, the point estimate corresponds to the DRM estimator, and the horizontal segment represents the bootstrap confidence interval. Table 7 reports estimates and 95% confidence intervals of the dominance indices.



(a) (b)

Figure 2: DRM estimates and bootstrap 95% CIs for various indicators.

Table 7 regards dominance indices, their DRM based point estimates and bootstrap CIs. The analysis reveals substantial economic heterogeneity across the six provinces. The two panels in Figure 2 provide a clear visualization of income levels and their associated uncertainty, highlighting pronounced disparities. Specifically, Figure 2b shows a distinct stratification in median incomes: Hubei and Hunan emerge as clear leaders, with estimated medians approaching 30,000 RMB, whereas

Sichuan (approximately 14,000 RMB) and Henan (16,362 RMB) occupy the lower end of the distribution. Fujian and Anhui fall between these two groups.

Table 7: Pairwise Dominant Indices $\gamma(F, G)$

Population F (Row)		Population G (Column)				
		Hubei	**Hunan**	**Fujian**	**Anhui**	**Sichuan**
Henan	Estimate	0.000	0.000	0.000	0.000	0.888
	CI	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.651)	(0.000, 0.000)	(0.289, 0.969)
Hubei	Estimate		0.290	0.991	0.995	0.998
	CI		(0.007, 0.988)	(0.930, 1.000)	(0.422, 0.998)	(0.989, 1.000)
Hunan	Estimate			0.969	0.926	0.993
	CI			(0.900, 0.997)	(0.724, 0.996)	(0.981, 0.998)
Fujian	Estimate				0.011	0.997
	CI				(0.000, 0.199)	(0.930, 0.999)
Anhui	Estimate					0.998
	CI					(0.978, 1.000)

Each cell contains the point estimate for $\gamma(F, G)$ (row F, column G) with its corresponding confidence interval (CI) in parentheses.

Figure 2a indicates that poverty rates in 2020 are generally below 8%, with Hubei and Hunan exhibiting comparatively lower levels.

The dominance index $\gamma(F_r, F_s)$ reported in Table 7 provides a global comparison of entire income distributions. The results reveal several strong dominance relationships. For example, the estimated dominance index between Hubei and Henan is 0.998, with a 95% confidence interval for $\gamma(F_r, F_s)$ given by $(0.989, 1.000)$, indicating that the income distribution of Hubei overwhelmingly dominates that of Henan. Similarly, Sichuan appears to be dominated by most other provinces, with weaker evidence relative to Henan.

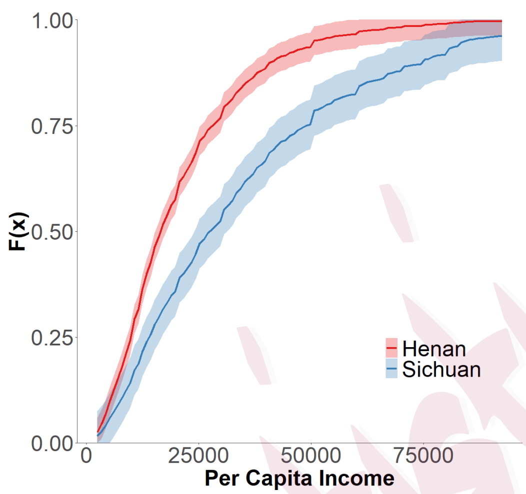
The dominance indices also uncover more nuanced relationships. For instance, $\widehat{\gamma}(\text{Fujian}, \text{Anhui}) = 0.011$ with confidence interval $(0.000, 0.199)$ suggests that, despite potential similarities in central tendencies, a larger proportion of residents in Anhui may have higher incomes than those in Fujian. Moreover, the confidence interval $(0.000, 0.651)$ for $\gamma(\text{Henan}, \text{Fujian})$ contains 0.5, indicating no statistically significant dominance between these two provinces.

To illustrate the implications of the process-level weak convergence established in Theorems 2 and 3, we construct 95% simultaneous confidence bands (SCBs) for the income distribution functions and quantile processes. Unlike pointwise confidence intervals, simultaneous confidence bands guarantee that the entire true function lies within the band with the prescribed coverage probability, providing a global assessment of distributional uncertainty.

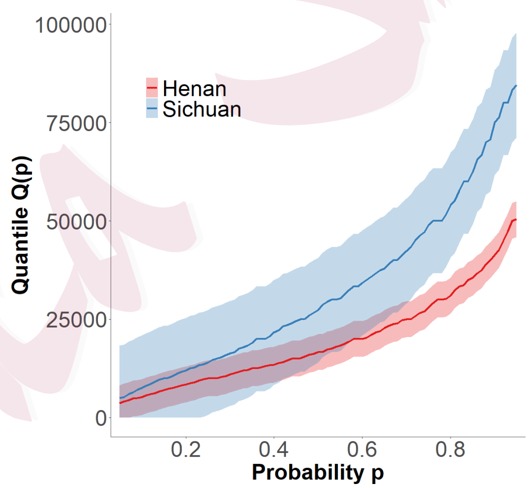
Based on our bootstrap framework, a $(1-\alpha)$ SCB for $F_r(x)$ can be constructed by finding the $(1-\alpha)$ empirical quantile, denoted as $q^*_{1-\alpha}$, of the bootstrap supremum statistic $\sup_x |\widehat{F}^*_r(x)-\widehat{F}_r(x)|$. The SCB is then given by $[\widehat{F}_r(x)-q^*_{1-\alpha}, \widehat{F}_r(x)+q^*_{1-\alpha}]$. A similar approach applies to the quantile process $Q_r(p)$.

For illustration, Figure 3 displays the DRM-fitted curves along with their corresponding 95% bootstrap SCBs for two selected provinces: Henan ($n = 1160$) and Sichuan ($n = 510$). Panel (a) compares their cumulative distribution functions (CDFs). Because Henan has a substantially larger sample size, its SCB is remarkably narrow. In contrast, the SCB for Sichuan is notably wider, reflecting the increased global uncertainty associated with a relatively smaller sample. Panel (b) compares the quantile processes of the same two provinces, where the progressive widening of the bands correctly captures the increasing uncertainty at the upper tails of the income distribution.

Beyond these sample size and tail effects, the SCBs provide profound insights into the global relationship between the two populations. Strikingly, in Panel (a), the entire SCB for Henan lies strictly above the SCB for Sichuan across the income domain (and conversely, Sichuan's quantile SCB lies strictly above Henan's in Panel (b)). This visualizes a strong first-order stochastic dominance, allowing us to rigorously conclude that the overall income level in Sichuan is significantly higher than that in Henan. Crucially, such global comparisons across the entire population are only made possible by our simultaneous inference framework. While previous pointwise convergence results could only infer differences at specific, pre-determined income thresholds or quantiles, the proposed SCBsgrounded in the weak convergence theoremsempower practitioners to draw comprehensive and definitive conclusions about the entire distribution at once.



(a)



(b)

Figure 3: DRM-estimated functions and their 95% bootstrap SCBs. Left panel (a): empirical processes for Henan and Sichuan. Right panel (b): quantile processes for Henan and Sichuan.

7. Conclusions

We develop a unified and comprehensive theory for bootstrap inference under the DRM. First, we establish distributional equivalence between DRM estimators and their bootstrap counterparts for model parameters. Second, we elevate the convergence of DRM distribution estimators from the pointwise level to weak convergence at the process level, thereby enabling rigorous asymptotic analysis for a broad class of statistical functionals. Third, we prove that the associated bootstrap processes converge to the same weak limits as their population analogues.

These theoretical results have substantive methodological implications. The bootstrap framework provides a principled and broadly applicable alternative to deriving and estimating analytically intricate covariance structures, which can be technically demanding under the DRM. This advantage is particularly pronounced for non-smooth functionals, such as quantiles and dominance indices, where direct variance estimation is often delicate. Moreover, process-level weak convergence furnishes a rigorous basis for simultaneous inference. As illustrated by the construction of simultaneous confidence bands in the real-data analysis, the proposed methodology enables global distributional comparisons that extend beyond pointwise assessments.

Taken together with the simulation studies, our results establish a rigorous and versatile foundation for bootstrap-based inference in DRM applications.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 72571262) and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2025-03989). The authors also acknowledge the computing resources provided by the University of British Columbia and the Digital Research Alliance of Canada.

Supplementary Materials

The online supplementary material contains detailed proofs of the theoretical results (Theorems 1–5 and auxiliary lemmas) established in Sections 3 and 4. It also provides additional descriptive statistics and empirical distribution plots for the real-data analysis presented in Section 6.

References

Anderson, J. (1979). Multivariate logistic compounds. *Biometrika 66*(1), 17–26.

Bondell, H. D. (2007). Testing goodness-of-fit in logistic case-control studies. *Biometrika 94*(2), 487–495.

Cai, S., J. Chen, and J. V. Zidek (2017). Hypothesis testing in the presence of multiple samples under density ratio models. *Statistica Sinica 27*(2), 761–783.

Chen, J. and Y. Liu (2013). Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics 41*, 1669–1692.

Chen, J. and J. Qin (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika 80*, 107–116.

Cheng, G. and J. Z. Huang (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics 38*(5), 2884.

Chernozhuokov, V., D. Chetverikov, K. Kato, and Y. Koike (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *The Annals of Statistics 50*(5), 2562–2586.

Davison, A. and D. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge, U.K.: Cambridge University Press.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics 7*, 1–26.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fokianos, K., B. Kedem, J. Qin, and D. A. Short (2001). A semiparametric approach to the one-way layout. *Technometrics 43*(1), 56–65.

Horowitz, J. L. and A. Krishnamurthy (2018). A bootstrap method for constructing pointwise and uniform confidence bands for conditional quantile funtions. *Statistica Sinica 28*(4), 2609–2632.

Hu, B. (2018). *An R package for monitoring test under density ratio model and its applications*. Ph. D. thesis, University of British Columbia.

Huang, A. and P. J. Rathouz (2012). Proportional likelihood ratio models for mean regression. *Biometrika 99*(1), 223–229.

Kaji, T. (2018). *Essays on Asymptotic Methods in Econometrics*. Ph. D. thesis, Massachusetts Institute of Technology.

Kedem, B., G. Lu, R. Wei, and P. D. Williams (2008). Forecasting mortality rates via density ratio modeling. *Canadian Journal of Statistics 36*(2), 193–206.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics. New York: Springer.

Li, P., Y. Liu, and J. Qin (2017). Semiparametric inference in a genetic mixture model. *Journal of the American Statistical Association 112*(519), 1250–1260.

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

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics 19*, 1725–1747.

Parente, P. M. and R. J. Smith (2014). Recent developments in empirical likelihood and related methods. *Annu. Rev. Econ. 6*(1), 77–102.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika 85*(3), 619–630.

Qin, J. (2017). *Biased Sampling, Over-identified Parameter Problems and Beyond.* New York: Springer.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Qin, J. and B. Zhang (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika 84*(3), 609–618.

Shao, U. and D. Tu (1995). *The Jackknife and Bootstrap.* New York: Springer-Verlag.

Soukarieh, I. and S. Bouzebda (2023). Renewal type bootstrap for increasing degree u-process of a markov chain. *Journal of Multivariate Analysis 195*, 105143.

Tan, Z. (2009). A note on profile likelihood for exponential tilt mixture models. *Biometrika 96*(1), 229–236.

Wu, C. and W. W. Lu (2016). Calibration weighting methods for complex surveys. *International Statistical Review 84*(1), 79–98.

Wu, C. and R. R. Sitter (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association 96*(453), 185–193.

Zhang, A. G. and J. Chen (2022). Density ratio model with data-adaptive basis function. *Journal of Multivariate Analysis 191*, 105043.

Zhuang, W., B. Hu, and R. Chen (2019). Semiparametric inference for the dominance index under the density ratio model. *Biometrika 106*(1), 229–241.

School of Management, University of Science and Technology of China, Hefei, Anhui, China

E-mail: weizh@ustc.edu.cn

School of Management, University of Science and Technology of China, Hefei, Anhui, China

E-mail: ywqyw123@mail.ustc.edu.cn

Department of Statistics, The University of British Columbia, Vancouver, BC, Canada

E-mail: jhchen@stat.ubc.ca