第51卷第12期 中国科学技术大学学报 Vol. 51, No. 12
2021年12月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Dec. 2021
Received: 2021-08-06; Revised: 2021-09-30 doi:10.52396/JUST-2021-0179

Statistical test for high order stochastic dominance under the density ratio model

ZHOU Yang[1], QIU Guoxin[2], ZHUANG Weiwei[3]*
1. School of Management, University of Science and Technology of China, Hefei 230026, China;
2. School of Business, Xinhua University of Anhui, Hefei 230088, China;
3. International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230061, China
* Corresponding author. E-mail: weizh@ustc.edu.cn

Abstract: In economics, medicine and other fields, how to compare the dominance relations between two distributions has been widely discussed. Usually population means or medians are compared. However, the population with a higher mean may not be what we will choose, since it may also have a larger variance. Stochastic dominance proposes a good solution to this problem. Subsequently, how to test stochastic dominance relations between two distributions is worth discussing. In this paper, we develop the test statistic of high order stochastic dominance under the density ratio model. In addition, we provide the asymptotic properties of test statistic and use the bootstrap method to obtain p-values to make decisions. Furthermore, the simulation results show that the proposed test statistics have the high test power.
Keywords: stochastic dominance; density ratio model; bootstrap test; empirical likelihood
CLC number: O212; O211.4; O211.6 **Document code**: A
2020 Mathematics Subject Classifications: 62E20; 60E15; 90C15

1 Introduction

Stochastic dominance (SD) is a method based on the expected utility theory, which can sort stochastic variables and reduce the effective set to help investors to make decisions. Stochastic dominance theory was originally proposed by Lehmann[1]. Later, Hadar and Russell[2], Hanoch and Levy[3], Rothschild and Stiglitz[4], and other scholars applied stochastic dominance theory and criteria to practice. Some scholars combined stochastic dominance with statistics to consider testing stochastic dominance. Regarding the test of stochastic dominance, it can be roughly divided into two categories: comparing the distributions at a finite number of grids, and comparing the distributions over the whole support. For the first type of test statistics, Anderson[5] established a test based on the t-statistic for two independent populations. Although their test statistic follows a normal distribution, the test power is not high. Davidson and Duclos[6] based on Anderson's test, used a new method to handle finite data. For the latter, McFadden[7] proposed the KS statistic to test first degree stochastic dominance, but the sample sizes of two populations are required to be equal. Eubank et al.[8] put forward a second degree stochastic dominance test, but the null hypothesis is that the distribution F dominates a known distribution F_0. In reality, both distributions may be unknown. Kaur et al.[9] advised to use the infimum of distributions to test the dominance relations. The advantage of this method is that the limiting distribution of test statistic can be given. However, the disadvantage is that if one distribution almost dominates the other one, the null hypothesis cannot be rejected. Schmid and Trede[10] proposed a test for second degree stochastic dominance and gave its critical value, but the test required one known distribution with the monotonically decreasing density. The strict assumption results in the narrow scope of application. Barrett and Donald[11] presented a test statistic based on KS and gave its asymptotic distribution. Donaid and Hsu[12] offered a method to improve the power of stochastic dominance test. Estimators chosen in these papers generally used empirical distributions. Such nonparametric methods sometimes may have large errors. Especially, in reality, the populations under comparison are usually of the same nature: In economics, they can be income distributions of several socio-demographic groups; in

Citation: Zhou Yang, Qiu Guoxin, Zhuang Weiwei. Statistical test for high order stochastic dominance under the density ratio model. J. Univ. Sci. Tech. China, 2021, 51(12): 868-878.

finance, they can be asset return distributions. In these cases, the density ratio model (DRM) provides a semiparametric model to connect these populations. When the density functions of populations meet certain assumptions, we can estimate each cumulative distribution function based on the pooled sample, therefore this model can improve estimation efficiency compared to that of nonparametric model[13-19].

The DRM originated from the logistic discriminant analysis of Anderson[20,21]. Anderson[13] formalized this model by setting the ratio of density functions of certain samples with similar information as the parameters family. Owen[22,23] proposed that the empirical likelihood can effectively handle the basic function in the DRM. Qin and Zhang[14] showed that the DRM could be used to solve the case-control logistic regression problem, estimated the parameters in the DRM using empirical likelihood, and finally gave a test to illustrate the feasibility. Qin[24] applied the DRM to the expected likelihood of case-control data. Keziou and Leoni-Aubin[15] formally equated the maximum empirical likelihood estimation of the parameters in the density ratio model with the maximum dual likelihood estimation. The dual empirical likelihood can be written as a specific expression, so it is more convenient to calculate and apply. Chen and Liu[16] gave the asymptotic distributions of quantile estimations based on the DRM. Zhuang et al.[25] estimated the relaxation indexes of stochastic dominance under the DRM. Compared with parametric models with the given distributions, the DRM can compensate for the loss of distribution errors and effectively reduce the risk of misprediction of the model distributions. Meanwhile, compared with nonparametric models, the DRM can make full use of similar information by making fewer assumptions, and improve the estimation accuracy.

In this paper, we use a semiparametric method to estimate the distribution functions F and G by using the empirical likelihood under the DRM. Based on the resulting estimators, we propose the test statistics of high order stochastic dominance, obtain the asymptotic properties of the test statistics, and construct the critical values. We conduct inferences of the test by p-value simulation using the bootstrap method. We select normal distributions and gamma distributions for the artificial data simulation, and use an actual example of stocks to illustrate the validity of our test. Simulation studies show that our test statistics substantially improve the estimation efficiency compared to the test statistics based on empirical distributions.

The rest of paper is organized as follows. In Section 2, a brief introduction of stochastic dominance and the DRM is given. In Section 3, our test statistics of high order stochastic dominance are proposed, the asymptotic properties of the test statistics are given, and a bootstrap method is developed to obtain p-values to make decisions. In Section 4, we apply our method to analyze two artificial examples and one actual example of stocks. Section 5 concludes the paper.

2 Notations and definitions

2.1 Stochastic dominance

Here are three commonly used dominance relations: first degree stochastic dominance (FSD), second degree stochastic dominance (SSD) and third degree stochastic dominance (TSD). As a simple example, if the return of the asset X in any case is higher than that of the asset Y, we will choose the asset X without hesitation. This is the simplest FSD relationship, but the conditions of FSD are too strict and hard to meet in daily life. Compared with FSD, SSD is more common, and it is aimed at the avaricious and risk averse people. TSD is aimed at the investors who are not only avaricious and risk averse, but also have diminishing levels of risk aversion. Let the cumulative distribution functions of random variables X and Y be F and G, respectively. Now we give the definition of stochastic dominance. Before giving the definition, we need to make the following assumptions:

Assumption 2.1 Assume that:

① F and G have common support $\mathscr{Z}=[\underline{z},\bar{z}]$, where $-\infty<\underline{z}<\bar{z}<\infty$;

② F and G are continuous functions on $\mathscr{Z}$, and $F(\underline{z})=G(\underline{z})=0$, $F(\bar{z})=G(\bar{z})=1$.

Assumption 2.1 is the general assumption in the literature, e.g. Barrett and Donald[11] and Linton et al.[26], which requires that F and G are both continuous in support $\mathscr{Z}$. The second part of Assumption 2.1 is not restrictive. If $G(z)=F(z)=0$ on an interval $[\underline{z},\eta]$, then $\mathscr{Z}$ can be defined as $[\eta,\bar{z}]$. Similarly, when $G(z)=F(z)=1$ on an interval $[\bar{z}-\eta,\bar{z}]$, we can define $\mathscr{Z}=[\underline{z},\bar{z}-\eta]$.

Next, we give the definition of stochastic dominance.

Definition 2.1[6] For $j\geqslant 1$, F is said to dominate G of order j, denoted by $F\succeq_j G$ $(X\succeq_j Y)$, if and only if

$$\mathscr{F}_j(z;F)\leqslant\mathscr{F}_j(z;G),\ \forall z\in\mathscr{Z} \tag{1}$$

where

$$\mathscr{F}_1(z;H)=H(z),$$
$$\mathscr{F}_j(z;H)=\int_{\underline{z}}^{z}\mathscr{F}_{j-1}(t;H)\,\mathrm{d}t,\ H=F,\ G.$$

In order to make the concept better understood, we give the following figures of FSD and SSD.

It can be seen from Figures 1 and 2 that the area of F below G is always more than the area F above G. It is worth mentioning that stochastic dominance of different

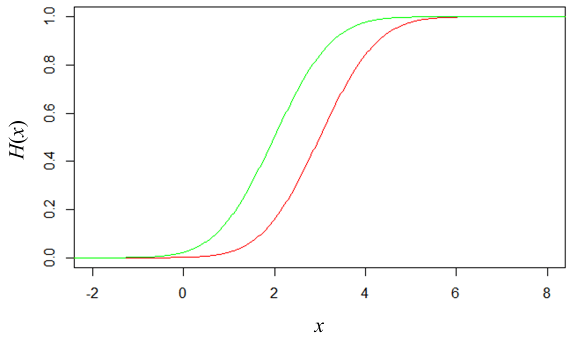

Figure 1. F (red line) and G (green line) satisfying FSD relationship.

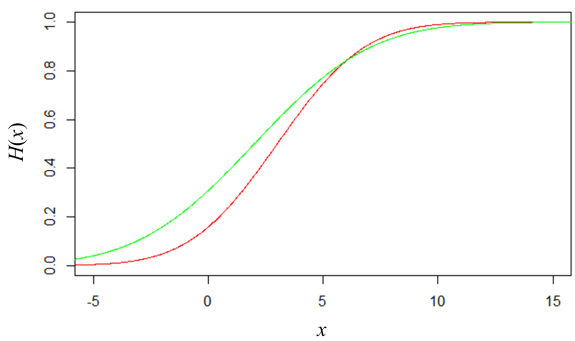

Figure 2. F (red line) and G (green line) satisfying SSD relationship.

orders satisfies the following relationship:

$$\text{FSD} \Rightarrow \text{SSD} \Rightarrow \text{TSD} \tag{2}$$

But the reverse is not true. The relationship (2) implies that we can test from low order to high order. If there is dominance relationship in the low order case, the high order dominance naturally exists. However, even if the higher order dominates, the lower order may not hold.

2.2 Density ratio model

In this subsection, we first briefly introduce the density ratio model. Suppose that F_0, F_1, $\cdots$, F_m, $m \geqslant 1$, are continuous cumulative distributions. These distributions are said to satisfy the density ratio model if they are linked through

$$\mathrm{d}F_k(x) = \exp\{\theta_k^{\mathrm{T}} q(x)\}\mathrm{d}F_0(x),\ k = 1,2,\cdots,m \tag{3}$$

where $q(x)$ is some pre-specified vector-values basis function and $\theta^{\mathrm{T}} = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}}, \cdots, \theta_m^{\mathrm{T}})$ is unknown parameter vector. We take $\theta_0 = 0$ for simplicity. Under the above assumption, these distributions have the same support. In this formulation, the baseline distribution F_0 is unspecified. The density ratio model is quite flexible and includes many common distribution families: the entire normal distribution family with $q(x) = (1, x, x^2)^{\mathrm{T}}$; the gamma distribution family with $q(x) = (1, x, \log x)^{\mathrm{T}}$. The components of $q(x)$ are linearly independent and its first element is one. The choice of $q(x)$ can be made case by case in applications. If the population distributions are normal-like, $q(x) = (1, x, x^2)^{\mathrm{T}}$ is a good choice, while for survival-type observations, $q(x) = (1, x, \log x)^{\mathrm{T}}$ is a good choice. It should be pointed out that the combination $q(x) = (1, x, x^2, \log|x|, \log^2|x|)^{\mathrm{T}}$ covers a mass of distribution families. Some papers have shown that for multi-sample situations with the same properties, the density ratio model has a good performance[14,16].

Given $k = 0, 1, \cdots, m$, suppose $n_k > 0$ is the number of observations from F_k, and x_{kj} represents the jth observation value from $F_k (j = 1, 2, \cdots, n_k)$. Given k, x_{k1}, x_{k2}, $\cdots$, x_{kn_k} are independent and identically distributed. The total number of observations is $n = n_0 + n_1 + \cdots + n_m$, and denote $\rho_k = n_k / n (k = 0, 1, \cdots, m)$ as the sample proportion. We first estimate the model parameters θ and F_0 through maximum likelihood estimation.

Let $p_{kj} = \mathrm{d}F_0(x_{kj})$. We have the log empirical likelihood function[27]:

$$l_n(\theta, p_{kj}) = \sum_{k,j} \log(p_{kj}) + \sum_{k,j} \theta_k^{\mathrm{T}} q(x_{kj}) \tag{4}$$

where the summation with respect to $\{k, j\}$ is over their entire ranges. The maximum empirical likelihood estimator $\hat{\theta}$ is the maximum point of (4). Keziou and Leoni-Aubin[15] indicated the equivalence of the maximum DRM-based empirical likelihood estimators and the dual maximum empirical likelihood estimators for both θ and F_0 in (3). Furthermore, Li et al.[28] carefully compared the DRM-based empirical likelihood and the dual empirical likelihood estimation methods under the two-sample density ratio model, and found that the two methods have the same point estimators for any underlying parameters. Otherwise, compared to the maximum DRM-based empirical likelihood method, the dual empirical likelihood estimation has a simpler analytical form and is easier to calculated. Therefore, in this paper, we take $\hat{\theta}$ as the maximum point of the following dual empirical likelihood function:

$$\tilde{l}_n(\theta, F_0) = \sum_{k,j} \log\{\sum_{r=0}^{m} \rho_r [\exp\{\theta_r^{\mathrm{T}} q(x_{kj})\}]\} + \sum_{k,j} \theta_k^{\mathrm{T}} q(x_{kj}) \tag{5}$$

Given the maximum dual EL estimator $\hat{\theta}$, the fitted values of p_{kj} are

$$\hat{p}_{kj} = \{nh(x_{kj}; \hat{\theta})\}^{-1},$$

where $h(x; \theta) = \sum_{k=0}^{m} \rho_k \exp\{\theta_k^{\mathrm{T}} q(x)\}$. Thus, the fitted population distribution F_k is given by

$$\hat{F}_k(x) = \sum_{r,j} \hat{p}_{rj} \exp\{\hat{\theta}_k^{\mathrm{T}} q(x)(x_{rj})\} \boldsymbol{I}(x_{rj} \leqslant x) =$$

$$n_k^{-1}\sum_{r,j} h_k(x_{rj};\hat\theta)\boldsymbol{I}(x_{rj}\leqslant x) \quad (6)$$

where $h_k(x;\theta)=\rho_k\exp\{\theta_k^{\mathrm T}q(x)\}/h(x;\theta)$, and $\boldsymbol{I}(A)$ denotes the indicator function of event A.

Assume F_k, $k=0,1,\cdots,m$, are continuous population distributions satisfying the DRM. Let $\hat F_k$ denote the corresponding estimator of F_k given by (6), for $k=0,1,\cdots,m$. For giving the asymptotic property of $n^{1/2}\{\hat F_k-F_k\}$, we need the following assumptions:

Assumption 2.2[16] Assume that

① the independent random samples $\{x_{kj}\}_{j=1}^{n_k}$ are from population F_k for $k=0,\cdots,m$;

② the total sample size $n=\sum_k n_k\to\infty$, and $\rho_k=n_k/n$ remains a constant;

③ the population distributions F_k satisfy the DRM with true parameter value θ^*, and $\int h(x;\theta)\,\mathrm dG_0<\infty$ in a neighborhood of θ^*;

④ the components of $q(x)$ are linearly independent and its first element is one.

Under Assumption 2.2, for any $0\leqslant r_1,r_2,\cdots,r_k\leqslant m$ and $x_1,x_2,\cdots,x_k$ in the support of F_0, $n^{1/2}\{\hat F_{r_j}(x_j)-F_{r_j}(x_j)\}$ are jointly asymptotically k-variate normal with mean $\mathbf 0$ and covariance matrix

$$\boldsymbol\Omega=(\omega_{r_i,r_j}(x_i,x_j))_{1\leqslant i\leqslant j}{}^{[16]}.$$

The analytical expression of $\boldsymbol\Omega$ is complex, and the (r,s)th entry of $\boldsymbol\Omega$ is determined as follows. Denote $x\wedge y=\min\{x,y\}$. Let $\delta_{rs}=1$ if $r=s$, and 0 otherwise. Let

$$\bar F(t)=\sum_r\rho_rF_r(t).$$

Then, the generic form of $\omega_{r_i,r_j}(x_i,x_j)$ is

$$\omega_{rs}(x,y)=\sigma_{rs}(x,y)-(\rho_r\rho_s)^{-1}\{a_{rs}(x\wedge y)-\boldsymbol B_r^{\mathrm T}(x)\boldsymbol W^{-1}\boldsymbol B_s(y)\} \quad (7)$$

where

$$\sigma_{rs}(x,y)=\rho_r^{-1}\delta_{rs}\{F_r(x\wedge y)-F_r(x)F_s(y)\},$$
$$a_{rs}(x)=\int_{-\infty}^{x}\{\delta_{rs}h_r(t)-h_r(t)h_s(t)\}\mathrm d\bar F(t),$$

and $\boldsymbol B_r(x)$ is a length-d vector, with its sth segment being

$$\boldsymbol B_{r,s}(x)=\int_{-\infty}^{x}\{\delta_{rs}h_r(t)-h_r(t)h_s(t)\}q(t)\mathrm d\bar F(t).$$

3 Test statistics and asymptotic properties

Select two different continuous distributions F_r and $F_s(r\neq s)$ from the distributions in the DRM. For the convenience of presentation, let F denote F_r, and let G denote F_s. To test the jth order dominance relations between F and G, $j\geqslant1$, we first formulate the null and alternative hypotheses as follows:

$\mathrm H_0^j$: $\mathscr F_j(z;F)\leqslant\mathscr F_j(z;G)$, for all $z\in[\underline z,\bar z]$;

$\mathrm H_1^j$: $\mathscr F_j(z;F)>\mathscr F_j(z;G)$, for some $z\in[\underline z,\bar z]$.

Barrett and Donald[11] and Donald and Hsu[12] considered the same hypotheses. They used empirical distributions to estimate the distribution functions F and G. Assume $X_1,X_2,\cdots,X_N$ and $Y_1,Y_2,\cdots,Y_M$ are independent and identically distributed samples from F and G, respectively. Their test statistics are defined as

$$\hat S_{\mathrm{EMP}}=\sqrt{\frac{NM}{N+M}}\sup_z\{\mathscr F_j(z;\hat F_N)-\mathscr F_j(z;\hat G_M)\},\ j\geqslant1 \quad (8)$$

where $\hat F_N(z)=\dfrac1N\sum_{i=1}^N\boldsymbol I(X_i\leqslant z)$, and

$$\hat G_M(z)=\frac1M\sum_{i=1}^M\boldsymbol I(Y_i\leqslant z).$$

However, in applied problems, the populations under comparison are generally of the same nature: In economics, they can be income distributions of several socio-demographic groups[5,6,29]; in finance, they are often asset return distributions[30,31]. In these cases, the density ratio model provides a semiparametric model to connect these populations. When the density functions of populations meet certain assumption, we can estimate each cumulative distribution function based on the pooled sample, therefore this model can improve estimation efficiency compared to that of nonparametric model[15-19].

We propose the semiparametric test statistics based on the density ratio model. Denote $\hat F_{\mathrm{DRM}}$ and $\hat G_{\mathrm{DRM}}$ the respective estimators of F and G given by (6). For $j\geqslant1$, we propose our test statistics as

$$\hat S_{\mathrm{DRM}}^j=\sqrt n\sup_{z\in\mathscr Z}\{\mathscr F_j(z;\hat F_{\mathrm{DRM}})-\mathscr F_j(z;\hat G_{\mathrm{DRM}})\} \quad (9)$$

In order to state the properties of test statistics, we first introduce the following lemma.

Lemma 3.1 On the common support set $\mathscr Z=[\underline z,\bar z]$ of F and G, for any $j\geqslant1$, we have

$$\mathscr F_j(x;F)=\frac{1}{(j-1)!}\int_{\underline z}^{x}(x-y)^{j-1}\mathrm dF(y),\ j\geqslant1 \quad (10)$$

Proof Since $F(\underline z)=0$, notice that

$$\mathscr F_1(x;F)=F(x),$$

and

$$\mathscr F_2(x;F)=\int_{\underline z}^{x}F(y)\mathrm dy=F(x)x-\int_{\underline z}^{x}y\,\mathrm dF(y)=\int_{\underline z}^{x}(x-y)\mathrm dF(y).$$

Suppose

$$\mathscr F_k(x;F)=\frac{1}{(k-1)!}\int_{\underline z}^{x}(x-y)^{k-1}\mathrm dF(y).$$

We want to prove

$$\mathscr F_{k+1}(x;F)=\frac{1}{k!}\int_{\underline z}^{x}(x-y)^{k}\mathrm dF(y).$$

Note that

$$\mathscr{T}_{k+1}(x;F)=\int_{\underline{z}}^{x}\mathscr{T}_k(y)\,dy=\frac{1}{(k-1)!}\int_{\underline{z}}^{x}\int_{\underline{z}}^{y}(y-s)^{k-1}dF(s)\,dy=\frac{1}{(k-1)!}\int_{\underline{z}}^{x}\int_{s}^{x}(y-s)^{k-1}dy\,dF(s)=\frac{1}{(k-1)!}\int_{\underline{z}}^{x}\frac{(x-s)^k}{k}dF(s)=\frac{1}{k!}\int_{\underline{z}}^{x}(x-y)^k\,dF(y).$$

By mathematical induction, the lemma is proved.

Next, we study the asymptotic properties of our test statistics. Before giving the theorem, we introduce the following notation. Let

$$W_n(x)=n^{1/2}[\{\widehat{F}_{\mathrm{DRM}}(x)-\widehat{G}_{\mathrm{DRM}}(x)\}-\{F(x)-G(x)\}],\ x\in[\underline{z},\bar{z}],$$

which converges weakly to a Gaussian process $W(x)$ in any finite dimensional distributions under Assumption 2.2[16]. The Gaussian process $W(x)$ has a continuous sample path, mean zero and covariance function

$$\mathrm{Cov}(W(x),W(y))=\omega_{rr}(x,x)+\omega_{ss}(y,y)-\omega_{rs}(x,y)-\omega_{sr}(y,x)\tag{11}$$

where the ω_{rs} are given in (7). When $x=y$, we obtain the variance function

$$\mathrm{Var}\{W(x)\}=\omega_{rr}(x,x)+\omega_{ss}(x,x)-2\omega_{rs}(x,x)\tag{12}$$

Theorem 3.1 Suppose that $F_0, F_1,\cdots, F_m$ are continuous population distributions satisfying the DRM. For any $0\leqslant r,\ s\leqslant m$, we denote $F_r=F$ and $F_s=G$. Let

$$\widehat{T}_j(\cdot)=\sqrt{n}\{[\mathscr{T}_j(\cdot;\widehat{F}_{\mathrm{DRM}})-\mathscr{T}_j(\cdot;\widehat{G}_{\mathrm{DRM}})]-[\mathscr{T}_j(\cdot;F)-\mathscr{T}_j(\cdot;G)]\},\ j\geqslant 1.$$

Then, under Assumption 2.2, we have

$$\widehat{T}_j(\cdot)\xrightarrow{\mathrm{w}}\mathscr{T}_j(\cdot;W),$$

where $\xrightarrow{\mathrm{w}}$ means converging in any finite dimensional distributions. Especially, when $j=1$, $\mathscr{T}_1(\cdot;W)$ is the Gaussian process with mean zero, and covariance function

$$\Omega_1(x,y,W)=\omega_{rr}(x,x)+\omega_{ss}(y,y)-\omega_{rs}(x,y)-\omega_{sr}(y,x).$$

Proof First, we need to show that for any $j\geqslant 1$,

$$\widehat{T}_j(\cdot)=\sqrt{n}\mathscr{T}_j(\cdot;\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\})\tag{13}$$

The case $j=1$ is obvious. For $j=2$, note the fact that

$$\mathscr{T}_2(z;F)=\int_{\underline{z}}^{z}F(t)\,dt=\int_{\underline{z}}^{z}\mathscr{T}_1(t;F)\,dt.$$

Consequently,

$$\mathscr{T}_2(z;\widehat{F}_{\mathrm{DRM}})-\mathscr{T}_2(z;F)-[\mathscr{T}_2(z;\widehat{G}_{\mathrm{DRM}})-\mathscr{T}_2(z;G)]=\int_{\underline{z}}^{z}\widehat{F}_{\mathrm{DRM}}(t)\,dt-\int_{\underline{z}}^{z}F(t)\,dt-[\int_{\underline{z}}^{z}\widehat{G}_{\mathrm{DRM}}(t)\,dt-\int_{\underline{z}}^{z}G(t)\,dt]=\int_{\underline{z}}^{z}\{[\widehat{F}_{\mathrm{DRM}}(t)-F(t)]-[\widehat{G}_{\mathrm{DRM}}(t)-G(t)]\}\,dt=\mathscr{T}_2(z;\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\}).$$

Now assume that when $j=k$, $k>2$, we have

$$\mathscr{T}_k(\cdot;\widehat{F}_{\mathrm{DRM}})-\mathscr{T}_k(\cdot;F)-[\mathscr{T}_k(\cdot;\widehat{G}_{\mathrm{DRM}})-\mathscr{T}_k(\cdot;G)]=\mathscr{T}_k(\cdot;\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\}).$$

Thus, when $j=k+1$,

$$\mathscr{T}_{k+1}(z;\widehat{F}_{\mathrm{DRM}})-\mathscr{T}_{k+1}(z;F)-[\mathscr{T}_{k+1}(z;\widehat{G}_{\mathrm{DRM}})-\mathscr{T}_{k+1}(z;G)]=\int_{\underline{z}}^{z}\{[\mathscr{T}_k(t;\widehat{F}_{\mathrm{DRM}})-\mathscr{T}_k(t;F)]-[\mathscr{T}_k(t;\widehat{G}_{\mathrm{DRM}})-\mathscr{T}_k(t;G)]\}\,dt=\int_{\underline{z}}^{z}\mathscr{T}_k(t;\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\})\,dt=\mathscr{T}_{k+1}(z;\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\}).$$

Therefore, (13) is proved.

Next, it is easy to see from (10) that, for any $j\geqslant 1$,

$$\widehat{T}_j(z)=\sqrt{n}\,\mathscr{T}_j(z;\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\})=\mathscr{T}_j(z;\sqrt{n}\{[\widehat{F}_{\mathrm{DRM}}-F]-[\widehat{G}_{\mathrm{DRM}}-G]\})=\mathscr{T}_j(z;W_n)\tag{14}$$

Under Assumption 2.2, it is known from Chen and Liu[16] that W_n weakly converges to the Gaussian process W in any finite dimensional distributions. Since $\mathscr{T}_j$ is a continuous function, it can be seen from Continuous Mapping Theorem[32] that $\mathscr{T}_j(z;W_n)$ weakly converges to $\mathscr{T}_j(z;W)$. Especially, when $j=1$, $\mathscr{T}_1(\cdot;W)=W$. From (11), we have known that, W is the Gaussian process with mean zero and covariance function

$$\Omega_1(x,y,W)=\omega_{rr}(x,x)+\omega_{ss}(y,y)-\omega_{rs}(x,y)-\omega_{sr}(y,x).$$

Then, this theorem is proved.

Finally, we construct the critical values for our test under the density ratio model, and discuss the asymptotic power properties of the test. For the null hypotheses H_0^j: $\mathscr{T}_j(z;F)\leqslant\mathscr{T}_j(z;G)$, $j\geqslant 1$, we use

$$\widehat{S}^j_{\mathrm{DRM}}=\sqrt{n}\sup_{z\in\mathscr{Z}}\{\mathscr{T}_j(z;\widehat{F}_{\mathrm{DRM}})-\mathscr{T}_j(z;\widehat{G}_{\mathrm{DRM}})\}$$

as the test statistics. We consider the test based on the decision rule of the form "reject H_0^j if $\widehat{S}^j_{\mathrm{DRM}}>c^j$", where c^j is some critical value. Let α be some desired significance level (say 0.05 or 0.01). We introduce the following notion:

$$\bar{S}_j=\sup_{z\in\mathscr{Z}}\mathscr{T}_j(z;W).$$

For any $j\geqslant 1$, we choose the critical values c^j satisfying

$$\mathbb{P}(\bar{S}_j > c^j) = \alpha.$$

The following theorem presents asymptotic power properties of our test.

Theorem 3.2 Suppose the same conditions as in Theorem 3.1, and c^j is a positive finite constant. Then, for $j \geqslant 1$, we have

(i) if H_0^j is true, then

$$\lim_{n\to\infty} \mathbb{P}(\text{reject } \mathrm{H}_0^j) \leqslant \mathbb{P}(\bar{S}_j > c^j) = \alpha \quad (15)$$

(ii) if H_1^j is true, then

$$\lim_{n\to\infty} \mathbb{P}(\text{reject } \mathrm{H}_0^j) = 1 \quad (16)$$

Proof For $j \geqslant 1$, under the condition that the null hypothesis H_0^j holds, we have

$$\widehat{S}^j_{\mathrm{DRM}} \leqslant \sup_{z\in\mathscr{Z}} \widehat{T}_j(z) + \sup_{z\in\mathscr{Z}} \sqrt{n}\{\mathscr{F}_j(z;F) - \mathscr{F}_j(z;G)\} \leqslant \sup_{z\in\mathscr{Z}} \widehat{T}_j(z).$$

From Theorem 3.1, it is known that $\widehat{T}_j(z) \xrightarrow{\mathrm{w}} \mathscr{F}_j(z;W)$.

Since $\bar{S}_j = \sup_z \mathscr{F}_j(z;W)$, we have

$$\lim_{n\to\infty} \mathbb{P}(\text{reject } \mathrm{H}_0^j) = \lim_{n\to\infty} \mathbb{P}(\widehat{S}^j_{\mathrm{DRM}} > c^j) \leqslant P(\bar{S}_j > c^j) = \alpha.$$

Part (i) is proved.

Next, we prove the second part. If the alternative hypothesis is true, there exists some $z^* \in \mathscr{Z}$ such that

$$\mathscr{F}_j(z^*;F) - \mathscr{F}_j(z^*;G) = \delta > 0.$$

Thus,

$$\widehat{S}^j_{\mathrm{DRM}} = \sup_{z\in\mathscr{Z}}\{\widehat{T}_j(z) + \sqrt{n}[\mathscr{F}_j(z;F) - \mathscr{F}_j(z;G)]\} \geqslant \widehat{T}_j(z^*) + \sqrt{n}\delta,\ j \geqslant 1.$$

It follows from Bahadur representation of DRM-based estimator[16] that

$$\sup_{z\in\mathscr{Z}} |\widehat{F}_{\mathrm{DRM}}(z) - F(z)| = O_p(n^{-1/2}),$$

and

$$\sup_{z\in\mathscr{Z}} |\widehat{G}_{\mathrm{DRM}}(z) - G(z)| = O_p(n^{-1/2}).$$

From (14), note that

$$\widehat{T}_j(z^*) = \mathscr{F}_j(z^*;\sqrt{n}\{[\widehat{F}_{\mathrm{DRM}} - F] - [\widehat{G}_{\mathrm{DRM}} - G]\}),\ j \geqslant 1.$$

Hence, when n goes to infinity, $|\widehat{T}_j(z^*)| < \infty$ and $\sqrt{n}\delta$ tends to infinity. It yields that

$$\lim_{n\to\infty}\mathbb{P}(\text{reject } \mathrm{H}_0^j) = 1.$$

Then, this theorem is proved.

The inequalities in (i) imply that the tests will never reject more often than α. Moreover, the result in (ii) implies that the tests are capable of detecting any violation of the full set of implications of the null hypothesis.

4 Simulation results

Although in the previous section, we construct the critical value c^j for the test, the value of c^j is difficult to obtain since the complex form of $\mathscr{F}_j(z;W)$. In this section, we conduct inferences of the test by p-value simulation using the bootstrap method. First, we select normal distributions and gamma distributions for the artificial data simulation. Then, we use the actual example of stocks to illustrate the validity of our test.

4.1 Bootstrap hypothesis tests

Assume that $\{X_1, X_2, \cdots, X_N\}$ are independently and identically distributed samples from F, and $\{Y_1, Y_2, \cdots, Y_M\}$ are independently and identically distributed samples from G. Define the pooled samples as $\{X_1, X_2, \cdots, X_N; Y_1, Y_2, \cdots, Y_M\}$. The detailed steps of the bootstrap approach are as follows:

Step 1 Compute $\widehat{S}_{\mathrm{DRM}}$ from the original sample $\{X_1, X_2, \cdots, X_N; Y_1, Y_2, \cdots, Y_M\}$.

Step 2 Draw the samples $\{X_1^*, X_2^*, \cdots, X_N^*\}$ from the pooled sample $\{X_1, X_2, \cdots, X_N; Y_1, Y_2, \cdots, Y_M\}$ with replacement and draw another samples $\{Y_1^*, Y_2^*, \cdots, Y_M^*\}$ in the same way.

Step 3 Use the samples in Step 2 to compute $\widehat{S}^*_{\mathrm{DRM}}$. Repeat Step 2 K times to get $\{\widehat{S}^*_{\mathrm{DRM},1}, \widehat{S}^*_{\mathrm{DRM},2}, \cdots, \widehat{S}^*_{\mathrm{DRM},K}\}$.

Step 4 Get p-value from $\widehat{p} = k/(K+1)$, where k is the number of $\widehat{S}^*_{\mathrm{DRM},i} \geqslant \widehat{S}_{\mathrm{DRM}}$, $i = 1, 2, \cdots, K$. For a given significance level α, if $\widehat{p} < \alpha$, then reject the null hypothesis.

In addition, about the selection of critical value c^j, we can use the quantile instead. Sorting the K times bootstrap samples from small to large, we get

$$\widehat{S}^*_{\mathrm{DRM},(1)} \leqslant \widehat{S}^*_{\mathrm{DRM},(2)} \leqslant \cdots \leqslant \widehat{S}^*_{\mathrm{DRM},(K)}.$$

For the given significance level α, the $1-\alpha$ quantile of the statistic $\widehat{S}_{\mathrm{DRM}}$ can be estimated by $\widehat{S}^*_{\mathrm{DRM},([K*(1-\alpha)])}$, where [.] denotes the rounding function. We can also compare c^j with $\widehat{S}_{\mathrm{DRM}}$ for the statistical inference.

Repeat Steps 1−4 B times, and record the number of $\widehat{p} < \alpha$ as b. When the alternative hypothesis holds, the rejection rate is b/B, which is used as test power in the simulation of this paper.

4.2 Two sample normal distributions

In the simulation of normal distribution, we select some different parameter settings for comparison. Under the DRM model, we choose the basis function of normal distribution as $q(x) = (1, x, x^2)^{\mathrm{T}}$. We set the sample size to be $N = M = 200$. The numbers of repetitions are $K = 300$ and $B = 500$. Table 1 gives several different situations. F and G in Table 1 represent the choice of distributions, and H_0^j: $\mathscr{F}_j(z;F) \leqslant \mathscr{F}_j(z;G)$, for all $z \in [\underline{z}, \bar{z}]$, $j = 1, 2, 3$, represent the null hypothesises in the case of jth order; H_1^j: $\mathscr{F}_j(z;F) > \mathscr{F}_j(z;G)$, for some $z \in [\underline{z}, \bar{z}]$, $j = 1, 2, 3$, are the corresponding alternative hypothesises for jth order.

Table 1. The dominance relationship of two normal distributions.

Design	F	G	$H_0^{(1)}$	$H_1^{(1)}$	$H_0^{(2)}$	$H_1^{(2)}$	$H_0^{(3)}$	$H_1^{(3)}$
1a	$N(2,3^2)$	$N(2,3^2)$	√		√		√	
1b	$N(2,3^2)$	$N(2,4^2)$		√		√	√	
1c	$N(2,3^2)$	$N(3,5^2)$		√		√		√

[Note] √ means that the hypothesis is valid.

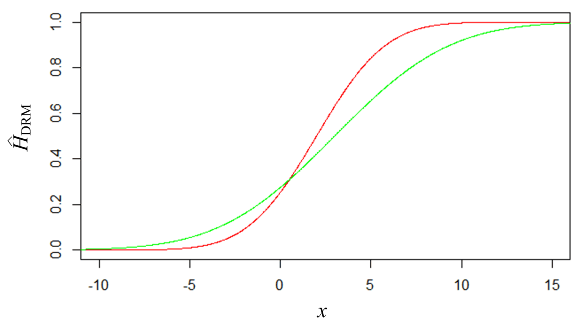

Figure 3. Estimation of distribution functions F (red line) and G (green line) under the density ratio model.

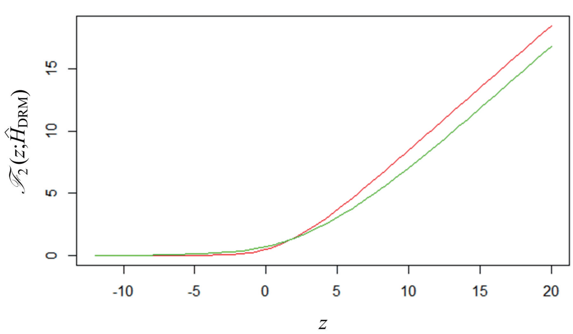

Figure 4. Estimation of functions $\mathscr{F}_2(z;F)$ (red line) and $\mathscr{F}_2(z;G)$ (green line) under the density ratio model.

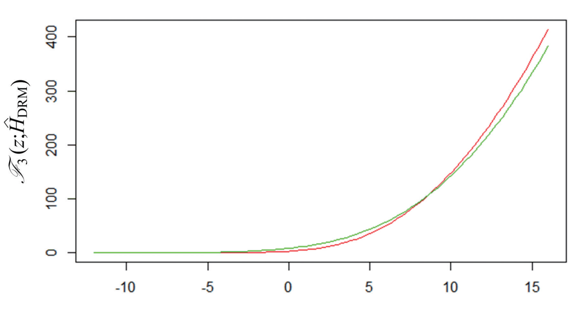

Figure 5. Estimation of functions $\mathscr{F}_3(z;F)$ (red line) and $\mathscr{F}_3(z;G)$ (green line) under the density ratio model.

In the special case 1a, the two distributions are always the same, no matter what value of j. For the remaining cases 1b and 1c, we will infer the dominance relations based on the distribution plots, and then give the rejection rates of the tests. We take the case 1c as the example. First, we present the distribution function plots of 1c for F and G under the density ratio model, as shown in Figure 3. From Figure 3 we can see that there is a crossover between the two distributions, so there is no first order stochastic dominance between F and G. Secondly, we want to judge whether there is second order stochastic dominance between the two distributions. We give the plots of

$$\mathscr{F}_2(z;F) = \int_{\underline{z}}^{z} F(t)\,\mathrm{d}t$$

and

$$\mathscr{F}_2(z;G) = \int_{\underline{z}}^{z} G(t)\,\mathrm{d}t,$$

as shown in Figure 4. From Figure 4, we can see that the two functions still have intersecting parts. F and G should not have second order stochastic dominance relationship.

Finally, we discuss whether there exists third order stochastic dominance. Figure 5 shows the plots of

$$\mathscr{F}_3(z;F) = \int_{\underline{z}}^{z} \mathscr{F}_2(t;F)\,\mathrm{d}t$$

and

$$\mathscr{F}_3(z;G) = \int_{\underline{z}}^{z} \mathscr{F}_2(t;G)\,\mathrm{d}t.$$

From the figure, we find that there is a crossover between the two functions, so there is still no third order stochastic dominance relationship.

Now we give the rejection rates under different orders. EMP represents the test statistic under empirical distributions; DRM represents the test statistic under the density ratio model. The numbers $j = 1, 2, 3$ in parentheses in the table indicate the jth order dominance. From Table 2, we can see that in the case of rejecting the null hypothesis, the rejection rates under the density ratio model are greater than those under empirical distributions. In the case of accepting the null hypothesis, the rejection rate should tend to 0, and the rejection rates under the density ratio model are smaller than those under empirical distributions. These all indicate that our test statistics under the density ratio model have better performance than the test statistic under empirical distributions.

4.3 Two sample gamma distributions

The DRM not only performs well in the case of common normal distributions, but also in the case of gamma distributions. We generate data from gamma distributions with the density function

$$f(x;\alpha,\beta)=\beta^{\alpha}x^{\alpha-1}\exp(-\beta x)/\Gamma(\alpha),\ x>0,$$

where α is the shape parameter and β is the scale parameter. For the same null and alternative hypothesises as Subsection 4.2, the sample sizes are $N=M=200$. Different from the normal distributions, the basis function for gamma distributions is chosen to be $q(x)=\{1,x,\log(x)\}^{\mathrm{T}}$. The numbers of repetitions are $K=300$ and $B=500$. The symbols in Table 3 have the same meanings as those in Table 1. Next, we take the case $2b$ as the example. It can be seen from Figure 6 that there is an intersection between the two distributions, and there is no phenomenon that one distribution is always above the other distribution. Hence, there is no first order stochastic dominance.

Table 2. Rejection rate of two normal distribution, $\alpha=0.05$.

Design	EMP (1)	DRM (1)	EMP (2)	DRM (2)	EMP (3)	DRM (3)
$1a$	0.086	0.052	0.068	0.060	0.052	0.048
$1b$	0.74	0.82	0.34	0.42	0.04	0
$1c$	0.84	0.92	0.518	0.622	0.14	0.37

Table 3. The dominance relationship of two gamma distributions.

Design	F	G	$\mathrm{H}_0^{(1)}$	$\mathrm{H}_1^{(1)}$	$\mathrm{H}_0^{(2)}$	$\mathrm{H}_1^{(2)}$	$\mathrm{H}_0^{(3)}$	$\mathrm{H}_1^{(3)}$
$2a$	$Ga(9,6)$	$Ga(9,6)$	√		√		√	
$2b$	$Ga(9,6)$	$Ga(2,1)$		√		√		√
$2c$	$Ga(9,4)$	$Ga(2,1)$		√	√		√	

[Note] √ means that the hypothesis is valid.

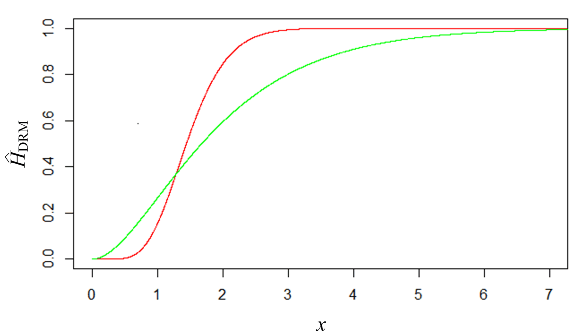

Figure 6. Estimation of distribution functions F (red line) and G (green line) under the density ratio model.

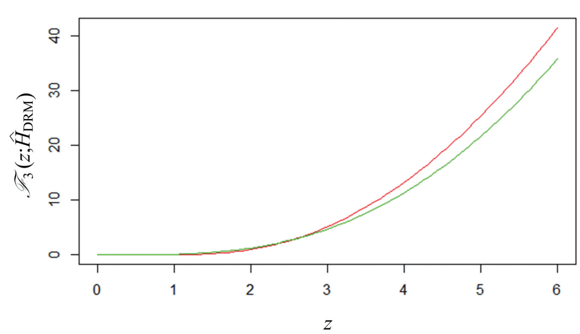

Figure 8. Estimation of functions $\mathscr{T}_3(z;F)$ (red line) and $\mathscr{T}_3(z;G)$ (green line) under the density ratio model.

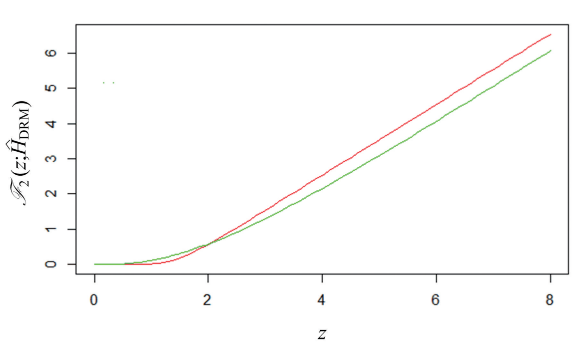

Figure 7. Estimation of functions $\mathscr{T}_2(z;F)$ (red line) and $\mathscr{T}_2(z;G)$ (green line) under the density ratio model.

In Figure 7, there is a partial region at the left end, where the function $\mathscr{T}_2(z;F)$ is above $\mathscr{T}_2(z;G)$. Therefore, we can infer that there is no second order stochastic dominance between the two distributions.

Figure 8 shows that the plots of

$$\mathscr{T}_3(z;F)=\int_{\underline{z}}^{z}\mathscr{T}_2(t;\ F)\,\mathrm{d}t$$

and

$$\mathscr{T}_3(z;G)=\int_{\underline{z}}^{z}\mathscr{T}_2(t;\ G)\,\mathrm{d}t$$

still have an intersection area, so there is no third order stochastic dominant relationship.

From Table 4, we can see that the rejection rates under empirical distributions are very close to those under the density ratio model. However, when rejecting the null hypothesis, the rejection rates under the density ratio model are slightly larger; when accepting the null hypothesis, the rejection rates under the density ratio model are slightly smaller. This phenomenon shows that our test statistics under the density ratio model are relatively more effective than the test statistics under empirical distributions.

Table 4. Rejection rate of two gamma distribution, $\alpha=0.05$.

Design	EMP(1)	DRM(1)	EMP(2)	DRM(2)	EMP(3)	DRM(3)
$2a$	0.06	0.052	0.054	0.048	0.056	0.052
$2b$	0.99	0.996	0.92	0.98	0.87	0.90
$2c$	0.89	0.96	0	0	0	0

Figure 9. Estimated distribution functions of DVN (dashed) and NOV (solid) under the density ratio model.

Figure 10. Second order function estimates for DVN (dashed) and NOV (solid) under the density ratio model.

4.4 Real data example

Finally, we will apply our method to the real data example of stocks. In recent years, the topic of stocks has become more and more popular. Using the stochastic dominance method to compare the strength and weakness of two stocks is an effective method. We select two stocks from Devon Energy Corporation (DVN) and National Oilwell Varco (NOV) over the last three years (2017−2019) for analysis. The reason for choosing the two stocks is that their closing prices are not much different, and it is difficult to determine the pros and cons. We scale the data due to the large magnitude of the data in the likelihood estimation. The null hypothesis is taken as DVN $\geq$ NOV. Sample sizes are $N = M = 753$, and the number of repetitions is $K = 300$. The basis function of the density ratio model is selected as $q(x) = (1, x, x^2)^{\mathrm{T}}$. Now, we give the distribution function plots of two stocks under the density ratio model.

From Figure 9, we can see that there is no first order stochastic dominance between the two stocks, but there may exist second order stochastic dominance. Next, we give the second order function plots of two stocks under the density ratio model.

It can be seen from Figure 10 that the second order function plot of DVN is always below that of NOV, which means that in the second order case, DVN dominates NOV. In addition, the value of the test statistic in second order case is -4.62×10^{-4}, which falls into the 95% confidence interval $(-\infty, 4.17 \times 10^{-2}]$, and the p-value is 0.52. These all show that we cannot reject the null hypothesis, so we can infer that the stock DVN is second stochastic dominant the stock NOV. For risk averse people, we would recommend the stock DVN.

5 Conclusions

In this paper, we propose a semiparametric method to test high order stochastic dominance relations between two different populations. We introduce the test statistics based on the DRM and prove their asymptotic properties. A bootstrap method is developed to obtain p-values for making decisions. The normal distributions and gamma distributions are selected for artificial data simulation. Simulation studies show that our test statistic substantially improves the estimation efficiency compared to the test statistic based on empirical distributions. Finally, we apply our method to an actual example of stocks. A topic for further work is the extension of our method to test almost stochastic dominance relations. Another possible application of the current inference framework is to test factional stochastic dominance, for example, stochastic dominance of order $1+\gamma$, for $0 < \gamma < 1$, which is related to stochastic optimization.

Acknowledgments

The work is supported by the National Nature Science Foundation of China (Nos. 71971204, 71871208, 11701518), and the Provincial Natural Science Foundation of Anhui (No. 1908085MG236).

Conflict of interest

The authors declare no conflict of interest.

Author information

ZHOU Yang is currently pursuing the master degree in Statistics and Finance with the School of Management, University of Science and Technology of China. His research interests include stochastic dominance and statistical test.

ZHUANG Weiwei (corresponding author) received the PhD degree in Probability and Statistics from the University of Science and Technology of China (USTC) in 2006. She is currently an Associate Professor with the Department of Statistics and Finance, USTC. Her research interests include statistical dependence, stochastic comparisons, semiparametric model, and their applications.

References

[1] Lehmann E L. Testing Statistical Hypotheses. New York: Wiley, 1959.

[2] Hadar J, Russell W R. Rules for ordering uncertain

prospects. The American Economic Review, 1969, 59: 25-34.
[3] Hanoch G, Levy H. The efficiency analysis of choices involving risk. The Review of Economic Studies, 1969, 36: 335-346.
[4] Rothschild M, Stiglitz J E. Increasing risk: I. A definition. Journal of Economic Theory, 1970, 2: 225-243.
[5] Anderson G. Nonparametric tests for stochastic dominance. Econometrica, 1996, 64: 1183-1193.
[6] Davidson R, Duclos J Y. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica, 2000, 68: 1435-1464.
[7] McFadden D. Testing for Stochastic Dominance. New York: Springer, 1989.
[8] Eubank R, Schechtman E, Yitzhaki S. A test for 2nd order stochastic dominance. Communications in Statistics: Theory and Methods, 1993, 22: 1893-1905.
[9] Kaur A, Rao B L S, Singh H. Testing for 2nd-order stochastic dominance of 2 distributions. Econometric Theory, 1994, 10: 849-866.
[10] Schmid F, Trede M. A Kolmogorov-type test for second-order stochastic dominance. Statistics and Probability Letters, 1998, 37: 183-193.
[11] Barrett G F, Donald S G. Consistent tests for stochastic dominance. Econometrica, 2003, 71:71-104.
[12] Donald S G, Hsu Y C. Improving the power of tests of stochastic dominance. Econometric Reviews, 2016, 35: 553-585.
[13] Anderson J A. Multivariate logistic compounds. Biometrika, 1979, 66: 17-26.
[14] Qin J, Zhang B. A goodness-of-t test for logistic regression models based on case-control data. Biometrika, 1997, 84: 609-618.
[15] Keziou A, Leoni-Aubin S. On empirical likelihood for semiparametric two-sample density ratio models. Journal of Statistical Planning and Inference, 2008, 138: 915-928.
[16] Chen J, Liu Y. Quantile and quantile-function estimations under density ratio model. The Annals of Statistics, 2013, 41: 1669-1692.
[17] Qin J. Biased Sampling, Over-identified Parameter Problems and Beyond. New York: Springer, 2017.
[18] Wang C, Marriott P, Li P. Testing homogeneity for multiple nonnegative distributions with excess zero observations. Computational Statistics and Data Analysis, 2017, 114: 146-157.
[19] Wang C, Marriott P, Li P. Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations. Journal of Multivariate Analysis, 2018, 166: 182-197.
[20] Anderson J A. Separate sample logistic discrimination. Biometrika, 1972, 59: 19-35.
[21] Anderson J A. Diagnosis by logistic discriminant function: Further practical problems and results. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1974, 23: 397-404.
[22] Owen A B. Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 1988, 75: 237-249.
[23] Owen A B. Empirical likelihood ratio confidence regions. The Annals of Statistics, 1990, 18: 90-120.
[24] Qin J. Inferences for case-control and semiparametric two-sample density ratio models. Biometrika, 1998, 85: 619-630.
[25] Zhuang W, Hu B, Chen J. Semiparametric inference for the dominance index under the density ratio model. Biometrika, 2019, 106: 229-241.
[26] Linton O, Song K, Whang Y J. An improved bootstrap test of stochastic dominance. Journal of Econometrics, 2010, 154: 186-202.
[27] Owen A B. Empirical Likelihood. New York: Chapman and Hall, 2001.
[28] Li H, Liu Y, Liu Y, et al. Comparison of empirical likelihood and its dual likelihood under density ratio model. Journal of Nonparametric Statistics, 2018, 30(3): 581-597.
[29] Bishop J A, Chow K V, Formby J P. A stochastic dominance analysis of growth, recessions and the U. S. income distribution, 1967-1986. Southern Economic Journal, 1991, 57(4): 936-946.
[30] Meyer J. Further applications of stochastic dominance to mutual fund performance. Journal of Financial and Quantitative Analysis, 1977, 12: 235-242.
[31] Gasbarro D, Wong W K, Kenton-Zumwalt J. Stochastic dominance analysis of iShares. The European Journal of Finance, 2007, 13(1): 89-101.
[32] Mann H B, Wald A. On the statistical treatment of linear stochastic difference equations. Econometrica, 1943, 11(3/4): 173-220.
[33] Kay R, Little S. Transformations of the explanatory variables in the logistic regression model for binary data. Biometrika, 1987, 74: 495-501.
[34] Kudo A. A multivariate analogue of one-sided tests. Biometrika, 1963, 50: 403-418.
[35] Owen A. Empirical likelihood for linear models. The Annals of Statistics, 1991, 19: 1725-1747.
[36] Pollard D. Convergence of Stochastic Processes. New York: Springer, 1984.
[37] Randles R H, Wolfe D A. Introduction to the Theory of Nonparametric Statistics. New York: Wiley, 1979.
[38] Shorack G R, Wellner J A. Empirical Processes with Applications in Statistics. New York: Wiley, 1986.
[39] Van Der Vaart A W, Wellner J A. Weak Convergence and Empirical Processes with Applications to Statistics. New York: Springer, 1996.
[40] Wolak F A. Testing inequality constraints in linear econometric models. Journal of Econometrics, 1989, 41: 205-235.

密度比模型下高阶随机占优的假设检验

周杨[1], 邱国新[2],庄玮玮[3*]
1. 中国科学技术大学管理学院,安徽合肥 230026;
2. 安徽新华学院商学院,安徽合肥 230088;
3. 中国科学技术大学管理学院国际金融研究院,安徽合肥 230601
* 通讯作者. E-mail:weizh@ ustc. edu. cn

摘要: 在经济学、医学等领域,如何比较两个分布的占优关系一直是人们关注的话题. 通常会比较平均值或中位数. 然而,具有更高均值的总体可能并不是最优的选择,因为它也可能具有更大的方差. 随机占优为这个问题提供了一个很好的解决方案. 那么,如何检验两个分布之间的随机占优就值得讨论. 本文研究了密度比模型下高阶随机优势的检验统计量. 此外,给出了检验统计量的渐近性,并使用自助法获得 p 值从而做出决策. 模拟结果表明本文提出的检验统计量具有较高的功效.

关键词: 随机占优;密度比模型;自助法;经验似然