[DOI] http://dx.doi.org/10.21487/jrm.2021.7.6.2.29

【연구경향】

Kernel Density Estimation for Polarization Measure

Na Kyeong Lee*

ABSTRACT

In this study, I propose a new estimator for the DER index that is a general measure of polarization introduced by Duclos et al.(2004). The existing estimators for the DER index are based on the empirical distribution function and a Rosenblatt-Parzen density estimator. The empirical distribution function, however, suffers from lack of smoothness. Hence, I suggest new estimators for the DER index using a new class of nonparametric kernel density estimators provided by Mynbaev and Martins-Filho(2010). I show that my estimators for polarization measure are consistent and establish asymptotic normality with a rate of convergence for the distribution of $\sqrt{n}$. A small Monte Carlo study reveals that my estimator performs well relative to the existing estimator for the DER index in terms of bias and mean squared error.

Keywords: Polarization Measure, DER Index, Nonparametric Density

---

* Assistant Professor, Department of Economics, Seoul Women's University. This work was supported by a research grant from Seoul Women's University (2021-0141).

Ⅰ. Introduction

According to the OECD(OECD 2011), over the past two decades, income inequality as measured by Gini coefficient, has widened in the most OECD countries. In the U.S., for example, more than 40 percent of total income is owned by the wealthiest 10 percent of the population. Thus, there has been growing interest in measuring inequality and the consequence of unequal economic performance during last few decades(Esteban and Ray 1994; Wolfson 1994; Zhang and Kanbur 2001; Duclos et al. 2004; Anderson et al. 2009; Anderson 2011).

The most well-known and widely used measures of inequality are Gini, Atkinson, and General entropy measures. The existing standard measures, however, are not able to explain all characteristics of inequalities. Since these measures focus on the deviation from the global mean and assessing the expected divergence, clustering around local mean or disappearing middle class can be explained by the existing inequality measures(Wolfson 1994)). Unlike inequality measure, polarization is varied given the following results: (1) when a population becomes more dense around the global mean. (2) divergence between two population poles increases. Therefore, polarization measures have been suggested to explain these characteristics that inequality measures fail to capture.

Esteban and Ray(1994) measured polarization in one dimension designed for discrete random variables. Later, Duclos et al.(2004) proposed a polarization measure called the DER index, and an estimator for the continuous random variables. Then they derived the estimator for the DER index based on a Rosenblatt-Parzen density estimator and an empirical distribution function that suffers from the lack of smoothness.

In this paper, for the estimation of the DER index, I use a new nonparametric kernel based density estimator, $\hat{f}_k$ that was introduced by Mynbaev and Martins-Filho(2010). This is a class of density estimators $\hat{f}_k$ achieves bias reduction relative to the Rosenblatt-Parzen density estimator $\hat{f}$. In addition, instead of using the empirical distribution function, I obtain the estimator of the DER index based on integrating $\hat{f}_k$. Integrating $\hat{f}_k$ seems desirable for the following reasons: (i) $\hat{f}_k$ forms a general class of nonparametric density estimators. When $k=1$, the newly developed density estimator $\hat{f}_k$ coincides with the Rosenblatt-Parzen density estimator $\hat{f}$. (ii) mean squared error of the integrating density estimator $\hat{f}$ might be smaller than the estimation of the empirical distribution function. (iii) Azzalini(1981), Falk et al.(1985) and Martins-Filho and Yao(2008) show that the estimation of the empirical distribution has asymptotic the same mean and same variance as the integrating a Rosenblatt-Parzan density estimator.

Throughout this paper, I assume that the true density f belongs to a Besov Space $B^r_{\infty,q}$where $1 \le q \le \infty$ and $r > 0$. This assumption is desirable since l-times continuous differentiability and uniform boundedness of f is stronger than $f \in B^r_{\infty,q}$ where $l < r$, that is $C^l(\mathbb{R}) \subseteq B^r_{\infty,q}$, where $C^l(\mathbb{R})$ denotes the space of l-times differentiable and continuous and bounded function in $\mathbb{R}$.

The first contribution of this paper is to show that my class of estimator for the DER index $\hat{P}_\alpha(\hat{F}_k)$ attain consistency. The second contribution of this paper is to establish the asymptotic normality of $\hat{P}_\alpha(\hat{F}_k)$. Lastly, I conduct a Monte Carlo study to investigate the finite sample performance of my estimators $P_\alpha(\hat{F})$. Then I propose and compare it to that of the existing estimator $P_\alpha(\tilde{F})$, where $\tilde{F}$ is the empirical distribution function. The simulation results indicate the performance improvement measured by the bias, and the root mean squared error.

The remainder of the paper is organized as follows. Section II provides the estimation of Polarization measure. Section III contains a Monte Carlo study that implements the proposed estimator and compares the performance with that of estimator suggested by Duclos et al.(2004). Lastly, section IV provides a summary and conclusion.

Ⅱ. Estimation for Polarization Measure

A. Finite difference and Besov Spaces

In this section, I provide a class of density estimators $\{\hat{f}_k\}_{k=1,2,\cdots}$ and an associate class of distribution estimators $\{\hat{F}_k\}_{k=1,2,\cdots}$ using the family of kernels $\{M_k\}_{k=1,2,\cdots}$ introduced by Mynbaev and Martins-Filho(2010). A series of definitions is needed to support the construction of the class. Properties of nonparametric density estimators and the smoothed estimators for the distribution function are traditionally obtained by using assumptions on the smoothness of the underlying density and cumulative distribution function. Smoothness can be regulated by finite differences which can be defined as forward, backward or centered. Let $C_s^l = \frac{s!}{(s-l)!l!}$ for $l = 1, 2, \cdots, s$ and $s \in \mathbb{Z}_+$ be binomial coefficients. A s-th order forward difference is defined by

$$\widetilde{\Delta}_h^s f(x) = \sum_{j=0}^{s} (-1)^{s-j} C_s^j f(x+jh) \tag{1}$$

where $s = 1, 2, \cdots,$ for x and $h \in \mathbb{R}$.

When I consider forward even-order difference, (1) can be written as

$$\tilde{\Delta}_h^{2k} f(x) = \sum_{|s|=0}^{k} c_{k,s} f(x+kh+sh) \tag{2}$$

where $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$ for $s=-k,\cdots,k$ and $k \in \{1,2,\cdots\}$. It is easy to verify that for $s=2k$, $\tilde{\Delta}_h^{2k} f(x) = \sum_{j=0}^{2k}(-1)^{2k-j} C_{2k}^{j} f(x+jh) = \sum_{|s|=0}^{k}(-1)^{s+k} C_{2k}^{s+k} f(x+kh+sh)$.

Next, I introduce Besov spaces $\mathcal{B}_{p,q}^{r}(\mathbb{R})$ where $1 \le p \le \infty$, $1 \le q \le \infty$, $r > 0$, and the norm in $\mathcal{B}_{p,q}^{r}(\mathbb{R})$ is defined by $\| f \|_{\mathcal{B}_{p,q}^{r}(\mathbb{R})} = \| f \|_{b_{p,q}^{r}} + \| f \|_{p}$ where the first part $\| f \|_{b_{p,q}^{r}}$ characterizes smoothness of f and is given by

$$\| f \|_{b_{p,q}^{r}} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} |\tilde{\Delta}_h^{2k} f(x)|^{p} dx \right)^{\frac{1}{p}}}{|h|^{r}} \right]^{q} \frac{dh}{|h|} \right\}^{1/q}$$

for $k \in \mathbb{Z}_{+}$ satisfying $2k > r$ (Tribel 1985; Mynbaev and Martins-Filho 2015). When $p=\infty$ and/or $q=\infty$, the integral(s) is(are) replaced by supremum. $C^{0}(\mathbb{R})$ is defined as the collection of all real-valued, bounded and uniformly continuous functions in $\mathbb{R}$, equipped with the norm $\| f \|_{C^{0}} = \sup_{x\in\mathbb{R}} |f(x)|$ (Tribel 2010). A full description of the relationships between $C^{l}(\mathbb{R})$ and a Besov space $\mathcal{B}_{p,q}^{r}$ can be founded in Besov et al.(1978). Given

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right), \tag{3}$$

the bias of my proposed estimators $\tilde{m}_k$ is expressed in terms of higher order finite differences. Let

$\lambda_{k,s}=\frac{(-1)^{s+k}(k!)^2}{(k+s)!(k-s)!}$ where $s=1,2,\cdots,k$. sGiven $-\frac{c_{k,s}}{c_{k,0}}=-\frac{c_{k,-s}}{c_{k,0}}=\lambda_{k,s}, s=1,2,\ldots,k,$ I derive $M_k(x)=\sum_{s=1}^{k}\frac{\lambda_{k,s}}{s}\left(K\left(\frac{x}{s}\right)+K\left(-\frac{x}{s}\right)\right)$. Consequently, $M_k(x)=M_k(-x)$ for $x\in\mathbb{R}$, that is M_k is symmetric. Since the coefficients $c_{k,s}$ satisfy $\sum_{|s|=0}^{k}c_{k,s}=(1-1)^{2k}=0,$ the following equation is true.

$$-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}c_{k,s}=1 \quad or \quad \sum_{s=1}^{k}\lambda_{k,s}=\frac{1}{2} \tag{4}$$

Equation (4) and $\int K(\psi)d\psi=1$ imply following equations:

$$\int M_k(\psi)d\psi=\sum_{s=1}^{k}\frac{\lambda_{k,s}}{s}\left[\int K\left(\frac{\psi}{s}\right)d\psi+\int K\left(-\frac{\psi}{s}\right)d\psi\right]=1$$

for all k.

Let $\{X_t\}_{t=1}^{n}$ be a random sample from a population having a density $f(x)$ for $x\in\mathbb{R}$. Mynbaev and Martins-Filho(2010) defines a new family of density estimators indexed by k evaluated at $x\in\mathbb{R}$ based on the kernel M_k as follows,

$$\hat{f}_k(x)=\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{X_t-x}{h_n}\right) \tag{5}$$

where h_n is a bandwidth sequence tending to zero as $n\to\infty$. When $k=1$ and K is symmetric, the density estimator in (5) coincides with the Rosenblatt-Parzen density estimator. Since the kernel $M_k(x)$ is symmetric, by using forward even-order differences (2), for a function f I have

$$\Delta_h^{2k}f(x)=\sum_{s=-k}^{k}c_{k,s}f(x+sh)$$

for $h \in \mathbb{R}$. It is easy to verify that $\widetilde{\Delta}_h^{2k} f(x) = \Delta_h^{2k} f(x+kh)$(Mynbaev and Martins-Filho 2015). Hence, I use centered even-order difference for a smoothness characteristic, and I have

$$\| f \|_{b_{p,q}^r} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} |\Delta_h^{2k} f(x)|^p dx \right)^{\frac{1}{p}}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{\frac{1}{q}}.$$

B. Cumulative Distribution Function Estimation

Let F be the distribution function with density f. Now I list assumption that will be used throughout the study.

ASSUMPTION 1: $\{X_t\}_{t=1}^n$ is an IID sequence.

ASSUMPTION 2: $F \in \mathcal{B}_{\infty,q}^r$ and $D^{(1)}F \in \mathcal{B}_{\infty,q}^r$ with $r > 0$ and $1 \leq q \leq \infty$.

ASSUMPTION 3: $h_n > 0$ for all n, $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

ASSUMPTION 4: For all $x \in \mathbb{R}$,

(1) $K(x): \mathbb{R} \to \mathbb{R}$ is a measurable function.

(2) $\int K(x)dx = 1$

(3) $\int |K(x)|dx < \infty$

(4) $\int xK(x)dx = 0$

(5) $\sup_{x \in \mathbb{R}} |K(x)| < C < \infty$ where $C \in \mathbb{R}$

(6) K has a compact support.

Let define a family of estimators for the distribution F by integrating $\hat{f}_k$, as below.

$$\hat{F}_k(x) = \int_{-\infty}^{x} \hat{f}_k(v)dv = \int_{-\infty}^{x} \frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t - v}{h_n}\right)dv.$$

Theorem 1

Let Assumption 1-4 *hold. For any* $h_n > 0, x \in \mathbb{R}$ *and* $k = 1, 2, \cdots,$

(a) $|Bias\left(\hat{F}_k(x)\right) = -\frac{1}{c_{k0}}\int_{-\infty}^{\infty} K(\psi)\Delta^{2k}_{-h_n\psi}F(x)dx$

(b) $|Bias\left(\hat{F}_k(x)\right) \le O(h_n^r)$ where $r < 2k$.

(c) $\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\left|Bias\left(\hat{F}_k(x)\right)\right| = 0.$

(d) $\hat{F}_k(x) - F(x) \xrightarrow{p} 0$ as $n \to \infty$.

Theorem 1 (c) implies that $\hat{F}_k$ is asymptotically unbiased and (d) deals with the consistency of $\hat{F}_k$.

C. Estimation of Polarization Measure

Duclos et al.(2004) construct a class of polarization measures based on the alienation-identification framework. An individual's sense of identification depends on other individuals which are similar to the individual. When an individual is located at income x, the individual's sense of identification depends on the density $x, f(x)$. Alienation is measured by means of a distance $|x-y|$. The DER is given by

$$P_\alpha(F) = \int_{\mathbb{R}} f(y)^\alpha a(y)dF(y)$$

where f and F are density and distribution function for income and α is a parameter related to the importance of the identification factor and is defined by the user. The values of α are described by the degree of polarization sensitivity and the greater is its value, the greater is different from inequality measurement. If $\alpha = 0$, the polarization measure resembles the Gini coefficient. The identification effect is denoted by both $f(y)^{\alpha}$ and $a(y)$ represents the alienation effect, with $a(y) = \int |x - y| dF(x)$ for all x and $y \in \mathbb{R}$.

An estimator of $P_{\alpha}(F)$ is given by substituting $F(y)$ by $\hat{F}_k$ and $\hat{a}_k(y)$ where $\hat{a}_k(y) \int |x - y| d\hat{F}_k(y)$. Hence, a family of estimators for the DER index is denoted by

$$P_{\alpha}(\hat{F}_k) = \int \hat{f}_k(y)^{\alpha} \hat{a}_k(y) d\hat{F}_k(y)$$

First, I consider the family of estimators for the alienation $a(y)$.

$$\hat{a}_k(y) = \int |x - y| d\hat{F}_k(x) = \int x\hat{f}_k(x)dx - y + 2\hat{F}_k(y) - 2y\int_{-\infty}^{y} x\hat{f}_k(x)dx$$

Before providing Theorem 2, it is necessary to propose the following Assumption 5.

ASSUMPTION 5:

(a) $\int |\psi| K(\psi) d\psi < \infty$

(b) $E[|X_t|^2] < \infty$

Theorem 2: *Let* ASSUMPTION 1-5 *hold. For every* $y \in \mathbb{R}$ *and* $k = 1,2,\cdots$,

(a) $\lim_{n\to\infty} |Bias(\hat{a}_k(y))| = 0$ *and* $Var(\hat{a}_k(y)) \to 0$ *as* $n \to \infty$.

(b) $\hat{a}_k(x) - a(x) = o_p(1)$.

Consistently, $\hat{a}_k(y)$ is a consistent estimator for $a(y)$ for $y \in \mathbb{R}$.

An estimator of the DER index, $\hat{P}_\alpha(F_k)$ is given by substituting the distribution function $F(y)$ by $\hat{F}_k(y)$. By replacing $a(y)$ and $f(y)^\alpha$ with $\hat{a}_k(y)$ and $\hat{f}_k(y)$, the estimator of $\hat{P}_\alpha(F_k)$ is denoted by below.

$$\hat{P}_\alpha(F_k) = \int \hat{f}_k(y)^\alpha \hat{a}_k(y) d\hat{F}_k(y).$$

The following theorem shows that a class of estimators for the DER index is asymptotically unbiased and consistent.

Theorem 3: *Let* ASSUMPTION 1-5 *hold. For* $k = 1, 2, \cdots, I$ *have*

(a) $Bias\left(\hat{P}_\alpha(\hat{F}_k)\right) \rightarrow 0$ as $n \rightarrow \infty$.

(b) $\hat{P}_\alpha(\hat{F}_k) - P_\alpha(F) = o_p(1)$.

The next theorem gives asymptotic normality of $\hat{P}_\alpha(\hat{F}_k)$. At this stage, I use Liapounov's CLT.

Theorem 4: *Let* ASSUMPTION 1-5 *hold. For* $k = 1, 2, \cdots,$ *note that*

$$\sqrt{n}[\hat{P}_\alpha(F) - E[\hat{P}_\alpha(F)]] \xrightarrow{d} N(0, V) \;\; N(0, \;\; V)$$

where

$$V \equiv Var\Big(-\alpha X_1 \int f(y)^{1+\alpha} dy - 2X_1 \int y f(y)^{1+\alpha} dy + 2(1+\alpha) f(X_1)^\alpha (1 - X_1^2) \iint_{-\infty}^{\theta} M_k(\psi) d\psi M\_k(\theta) d\theta\Big)$$

Ⅲ. Monte Carlo Study

In this section, I perform Monte Carlo study to investigate the finite sample performance of the proposed estimators for polarization measure. For comparison purpose, I implement the existing estimators that is provided by Duclos et al.(2004), which is given by $\hat{P}_\alpha(\tilde{F}) = \int \hat{f}(y)^\alpha \tilde{a}(y) d\tilde{F}(y) = n^{-1}\sum_{i=1}^{n} \hat{f}(y_i)^\alpha \tilde{a}(y_i)$ for $y_1 \le y_2 \le \cdots \le y_n$. Note that y_i is the empirical quantile for percentiles between $(i-1)/n$ and i/n. $\tilde{a}(y_i)$ is defined as $\tilde{a}(y_i)$ is defined as $\tilde{a}(y_i) = \tilde{\mu} + y_i(n^{-1}(2i-1)-1) - n^{-1}\left(2\sum_{j=1}^{i-1} y_j + y_i\right)$, where $\tilde{\mu}$ is the sample mean. $\hat{f}(y_i)^\alpha$ is estimated using Rosenblatt-Parzen density estimator. I consider simulated data from two different densities. They are: 1) Bimodal $\left(f_1(x) = \frac{3}{4}N(-1, 1.8) + \frac{1}{4}N(4, 0.4)\right)$ and 2) Trimodal $\left(f_2(x) = \frac{1}{5}N(0, 0.5) + \frac{3}{5}N(5,1) + \frac{2}{5}N(7, 2.5)\right)$. For each these densities 1000 samples of size and $n = 200,\ 800$ were generated.

〈Table 1〉 Bias(B) and Root Mean Squared Error(R) for $f_1(x)$

$n = 200$	$\alpha = 0.25$		$\alpha = 0.50$		$\alpha = 0.75$		$\alpha = 1.0$	
	B	R	B	R	B	R	B	R
$\tilde{P}_\alpha(\tilde{F})$	0.0366	0.0679	0.0535	0.0623	0.0928	0.0945	0.0594	0.0605
$\hat{P}_\alpha(\hat{F}_1)$	0.0338	0.0735	0.0503	0.0632	0.0911	0.0937	0.0580	0.0596
$\hat{P}_\alpha(\hat{F}_2)$	0.0202	0.0685	0.0298	0.0503	0.0707	0.0747	0.0411	0.0441
$\hat{P}_\alpha(\hat{F}_3)$	0.0198	0.0689	0.0259	0.0490	0.0657	0.0704	0.0364	0.0402
$\hat{P}_\alpha(\hat{F}_4)$	0.0203	0.0693	0.0245	0.0487	0.0636	0.0686	0.0343	0.0385

$n = 800$	$\alpha = 0.25$		$\alpha = 0.50$		$\alpha = 0.75$		$\alpha = 1.0$	
	B	R	B	R	B	R	B	R
$\tilde{P}_\alpha(\tilde{F})$	0.0178	0.0363	0.0438	0.0487	0.0382	0.0411	0.0353	0.0370
$\hat{P}_\alpha(\hat{F}_1)$	0.0097	0.0341	0.0381	0.0441	0.0347	0.0381	0.0328	0.0348
$\hat{P}_\alpha(\hat{F}_2)$	0.0063	0.0341	0.0265	0.0355	0.0204	0.0269	0.0190	0.0233
$\hat{P}_\alpha(\hat{F}_3)$	0.0036	0.0340	0.0220	0.0326	0.0153	0.0237	0.0143	0.0201
$\hat{P}_\alpha(\hat{F}_4)$	0.0032	0.0341	0.0205	0.0318	0.0134	0.0228	0.0123	0.0190

〈Table 2〉 Bias(B) and Root Mean Squared Error(R) for $f_2(x)$

$n = 200$	$\alpha = 0.25$		$\alpha = 0.50$		$\alpha = 0.75$		$\alpha = 1.0$	
	B	R	B	R	B	R	B	R
$\tilde{P}_\alpha(\tilde{F})$	0.0462	0.0917	0.0420	0.0576	0.0643	0.0671	0.0671	0.0679
$\hat{P}_\alpha(\hat{F}_1)$	0.1282	0.1556	0.0942	0.1065	0.0989	0.1028	0.0899	0.0971
$\hat{P}_\alpha(\hat{F}_2)$	0.0743	0.1249	0.0503	0.0774	0.0637	0.0725	0.0625	0.0663
$\hat{P}_\alpha(\hat{F}_3)$	0.0513	0.1168	0.0319	0.0701	0.0493	0.0619	0.0515	0.0569
$\hat{P}_\alpha(\hat{F}_4)$	0.0390	0.1140	0.0218	0.0680	0.0414	0.0568	0.0455	0.0520

$n = 800$	$\alpha = 0.25$		$\alpha = 0.50$		$\alpha = 0.75$		$\alpha = 1.0$	
	B	R	B	R	B	R	B	R
$\tilde{P}_\alpha(\tilde{F})$	0.0162	0.0480	0.0189	0.0313	0.0247	0.0288	0.0205	0.0227
$\hat{P}_\alpha(\hat{F}_1)$	0.0191	0.0525	0.0208	0.0354	0.0266	0.0318	0.0217	0.0247
$\hat{P}_\alpha(\hat{F}_2)$	-0.0034	0.0501	-0.0032	0.0300	0.0041	0.0192	0.0024	0.0131
$\hat{P}_\alpha(\hat{F}_3)$	-0.0077	0.0508	-0.0087	0.0314	-0.0014	0.0192	-0.0026	0.0135
$\hat{P}_\alpha(\hat{F}_4)$	-0.0076	0.0509	-0.0097	0.0319	-0.0029	0.0195	-0.0041	0.0140

I observe the following general regularities. First, as predicted by the suggested asymptotic results, for all densities considered the bias and root mean squared error of the estimators $\hat{P}_\alpha(\hat{F}_k)$ for $k=1,\ 2,\ 3,\ 4$ fall as the sample size increases from 200 to 800. Second, the increase in the values of k reduces bias and root MSE, but this is not verified for all experiments. In Table 1, the estimator bias for $k=2,3,4$ is smaller than the existing estimator $\hat{P}_\alpha(\hat{F}_1)$ for $n=200, 800$. It is observed that Bias and RMSE fall when k increases except for $\alpha=0.25$. The results are consistent with the proposed theory and finding.

Similarly, in Table 2, the existing estimators outperform than $\hat{P}_\alpha(\hat{F}_1)$. However, $\hat{P}_\alpha(\hat{F}_2)$, $\hat{P}_\alpha(\hat{F}_3)$ and $\hat{P}_\alpha(\hat{F}_4)$ show better performance than the existing estimators for polarization measure in terms of Bias and RMSE. For $k=4$, the case where the smallest bias reductions are attained, bias is reduced by as much as 80 percent relative to the estimators $\tilde{P}_\alpha(\tilde{F})$. Additionally, the magnitude of bias reduction produced by the estimators increases with the sample size. Finally, density functions with large curvature(in increasing order of curvature f_1 and f_2) are more difficult to estimate both in terms of bias and root mean MSE for the estimators.

Ⅳ. Summary and Conclusion.

Duclos et al.(2004) introduced polarization index and its estimators. The estimators are derived by using a Rosenblatt-Parzen density estimator and the empirical distribution function. Since the empirical distribution function jumps up by $1/n$ at each of the n data points, the empirical distribution suffers from lack of smoothness. In this paper, for the estimation of polarization measure, I used a class of new kernels

$M_k(\cdot)$ and density estimators $\hat{f}_k$ introduced by Mynbaev and Martins-Filho(2010) then provided a class of density estimators $\hat{F}_k$. Then I provide a new class of estimators for the DER index that achieve consistency and establish asymptotic normality. Finally, a Monte Carlo simulation results show that the suggested estimators outperform the existing estimators in terms of bias and root mean squared errors.

In political science, there are scholarly debates related to whether the mass public is ideologically polarized as the elites are. That is, there is no consensus on the ideological mass polarization at the mass level. To measure the polarization, DiMaggio et al.(1996), Balanda and MacGillvray(1988), Mouw and Sobel(2001), Myers(2007), Downey and Huffman(2001) provide the results of a polarization given the distribution's moments. On the other hands, Lee(2015) finds that the time varying pattern of the mass polarization between 1984 to 2008 based on the kernel based relative distribution method. By using the suggested polarization measure, the exact mass polarization can be measured as well as the elite polarization.

V. References

Anderson, Gordon. 2011. "Polarization measurement and inference in many dimensions when subgroups can not be identified." *Economics* 5(1): 20110011.

Anderson, Gordon, Oliver B. Linton, and Yoon-Jae Whang. 2009. "Nonparametric estimation of a polarization measure." *LSE STICERD Research Paper No. EM534*. Available at SSRN: https://ssrn.com/abstract=1547618.

Azzalini, Adelchi. 1981. "A note on the estimation of a distribution function and quantiles by a kernel method." *Bimetrika* 68(1): 326-328.

Balanda, Kevin P. and H. L. MacGillivray. 1988. "Kurtosis: a critical review." *American Statistician* 42(2): 111-119

Besov, Oleg V., P. Il'in Valentin and Sergei M. Nikol'skii. 1978. *Integral Representations of Functions and Imbedding Theorems. Vol. 1.* John Wiley & Sons

DiMaggio, Paul, John Evans and Bethany Bryson. 1996. "Have American's social attitudes become more polarized?" *American Journal of Sociology* 102(3): 690-755

Downey, Dennis J. and Matt L. Huffman. 2001. "Attitudinal polarization and trimodal distributions: measurement problems and theoretical implications." *Social Science Quarterly* 82(3): 494-505

Duclos, Jean-Yves, Joan Esteban and Debraj Ray. 2004. "Polarization: concepts, measurement, estimation." *Econometrica* 72(6): 1737-1772

Esteban, Joan-Maria, and Debraj Ray. 1994. "On the measurement of polarization." *Econometrica: Journal of the Econometric Society* 62(4): 819-851

Falk, Michael. 1985. "Asymptotic normality of the kernel quantile estimator." *The Annals of Statistics*, 13(1): 428-433

Lee, Jae Mook. 2015. "Assessing mass opinion polarization in the U.S using relative distribution method." *Social Indicators Research* 124(2): 571-598

Martins-Filho, Carlos, and Feng Yao. 2008. "A smooth nonparametric conditional quantile frontier estimator." *Journal of Econometrics* 143(2): 317-333

Mouw, Ted, and Michael E. Sobel. 2001. "Culture wars and opinion polarization: the case of abortion." *American Journal of Sociology* 106(4): 913-943

Mynbaev, Kairat, and Carlos Martins-Filho. 2010. "Bias reduction in kernel density estimation via Lipschitz condition." *Journal of Nonparametric Statistics* 22(2):219-235

Mynbaev, Kairat, and Carlos Martins-Filho. 2015. "Consistency and asymptotic normality for a nonparametric prediction under measurement errors." *Journal of Multivariate Analysis* 139: 166-188

Myers, Charles D. 2007. "Campaign intensity and polarization." Unpublished manuscript.

OECD. 2011. *Divided we Stand: Why Inequality Keeps Rising.* OECD Publishing, Paris, https://doi.org/10.1787/9789264119536-en.

Tribel, H. 1985. *Theory of function spaces*. Birkhäuser.

Tribel, H. 1985. *Theory of function spaces II*. Birkhäuser.

Wolfson, Michael C. 1994. "When inequalities diverge." *The American Economic Review* 84(2): 353-358

Zhang, Xiaobo, and Ravi Kanbur. 2001. “What difference do polarization measures make? An application to China.” *Journal of development studies* 37(3): 85-98

Ⅵ. Appendix

Proof. Theorem 1

(a) Given the independent and identically distributed(IID) assumption (maintained everywhere)- and $M_k(-x) = M_k(x)$, I have

$$\hat{F}_k(x) = \int_{-\infty}^{x} \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{v - X_t}{h_n}\right) dv = \frac{1}{n} \sum_{t=1}^{n} \int_{-\infty}^{\frac{x-X_t}{h_n}} M_k(\psi)\, d\psi$$

$$= \frac{1}{n} \sum_{t=1}^{n} \int_{-\infty}^{\frac{x-X_t}{h_n}} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{\psi}{s}\right) d\psi$$

$$= \frac{1}{n} \sum_{t=1}^{n} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \mathcal{G}\left(\frac{x - X_t}{s h_n}\right)$$

where $\mathcal{G}(x) = \int_{-\infty}^{x} K(v) dv.$

$$E[\hat{F}_k(x)] = E\left[\frac{1}{n} \sum_{t=1}^{n} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \mathcal{G}\left(\frac{x - X_t}{s h_n}\right)\right]$$

$$= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \int c_{k,s} \mathcal{G}(\psi)\, dF(x - sh_n\psi)(-1)$$

$$= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \left[-[\mathcal{G}(\psi) F(x - sh_n\psi)]|_{\psi=-\infty}^{\psi=+\infty} + \int K(\psi) F(x - sh_n\psi)\, d\psi\right]$$

$$= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \int K(\psi) F(x - sh_n\psi)\, d\psi$$

Hence, I have the following bias for $\hat{F}_k$.

$$Bias\left(\hat{F}_k(x)\right) = E[\hat{F}_k(x) - F(x)] = -\frac{1}{c_{k,0}} \sum_{|s|=0}^{k} c_{k,s} \int K(\psi) F(x - sh_n\psi)\, d\psi$$
$$= -\frac{1}{c_{k,0}} \int K(\psi)\, \Delta_{-h_n\psi}^{2k} F(x)\, d\psi \tag{6}$$

(b) From the result of (a), I have

$$|Bias(\hat{F}_k(x))| = \left| -\frac{1}{c_{k,0}} \int K(\psi) \Delta^{2k}_{-h_n\psi} F(x) d\psi \right|$$

$$= \left| -\frac{1}{c_{k,0}} \right| \left[\int \{ |K(\psi)| |h_n \psi|^{r+1/q} \}^{q'} d\psi \right]^{1/q'} \left[\int \left\{ \frac{\sup_{x\in\mathbb{R}} |\Delta^{2k}_{-h_n\psi} F(x)|}{|h_n \psi|^{r+1/q}} \right\}^{q} d\psi \right]^{1/q}$$

$$= h_n^r \left| -\frac{1}{c_{k,0}} \right| \left[\int \{ |K(\psi)| |\psi|^{r+1/q} \}^{q'} d\psi \right]^{1/q'} \|F\|^r_{\infty,q}$$

where $1/q + 1/q' = 1$ for $1 \le q \le \infty$.

(c) The proof is trivial. Given the assumption $h_n \to 0$ as $n \to \infty$, I have (c).

(d)

$$Var\left(\hat{F}_k(x)\right)$$

$$= E\left[\left(\hat{F}_k(x)\right)^2\right] - \left(E[\hat{F}_k(x)]\right)^2$$

$$= \int \left[\frac{1}{n} \sum_{t=1}^{n} \int_{-\infty}^{\frac{x-X_t}{h_n}} M_k(\psi) d\psi \right]^2 f(X_t) dX_t - \left\{ \int \frac{1}{n} \sum_{t=1}^{n} \int_{-\infty}^{\frac{x-X_t}{h_n}} M_k(\psi) d\psi f(X_t) dX_t \right\}^2$$

$$= \frac{1}{n} \int \left[\int_{-\infty}^{\phi} M_k(\psi) d\psi \right]^2 f(x - h_n\phi) d\phi - \frac{1}{n} \left\{ \int \int_{-\infty}^{\phi} M_k(\psi) d\psi f(x - h_n\phi) h_n d\phi \right\}^2$$

$$= \frac{1}{n} \int \left[\int_{-\infty}^{\phi/s} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} K(v) dv \right]^2 f(x - h_n\phi) h_n d\phi$$
$$- \frac{1}{n} \left\{ \int \int_{-\infty}^{\phi/s} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} K(v) dv f(x - h_n\phi) h_n d\phi \right\}^2$$

$$= \frac{1}{n} \left[-\frac{1}{c_{k,0}} \right]^2 \int \left[\sum_{|s|=1}^{k} c_{k,s} \mathcal{G}\left(\frac{\phi}{s}\right) \right]^2 dF(x - h_n\phi)(-1) - \frac{1}{n} \left[-\frac{1}{c_{k,0}} \right]^2 \left\{ \sum_{|s|=1}^{k} c_{k,s} \int \mathcal{G}\left(\frac{\phi}{s}\right) dF(x - h_n\phi)(-1) \right\}^2$$

$$= \frac{1}{n} \left[-\frac{1}{c_{k,0}} \right]^2 \left\{ 2 \sum_{|s|=1}^{k} c_{k,s} \int \mathcal{G}(\psi) K(\psi) F(x - h_n\psi) d\psi - \left[\sum_{|s|=1}^{k} c_{k,s} \int K(\psi) F(x - s h_n \psi) d\psi \right]^2 \right\}$$

$$= O(n^{-1})$$

since $\sup_{x\in\mathbb{R}} |F(x)| \le 1$ and $|\mathcal{G}(\phi)| = \left| \int_{-\infty}^{\phi} K(v) dv \right| \le C$, K has a compact support. Therefore, as $n \to \infty$, I have $Var\left(\hat{F}_k(x)\right) \to 0$ and $|Bias(\hat{F}_k(x))| \to 0$.

Proof of Theorem 2

$$E[\hat{a}_k(y)] - a(y)$$

$$= E\left[\int y\hat{f}_k(y)\,dy - \mu\right] + 2E\left[\hat{F}_k(y) - F(y)\right] - 2yE\left[\int_{-\infty}^{y} x\hat{f}_k(x)\,dx - \int_{-\infty}^{y} xf(x)\,dx\right]$$

$$= E\left[\int y\hat{f}_k(y)\,dy - \mu\right] + 2E\left[\hat{F}_k(y) - F(y)\right] - 2y\left[\int_{-\infty}^{y} xE\left[\hat{f}_k(x) - f(x)\right]dx\right]$$

Given the assumption $\int \psi K(\psi) = 0$,

$$E\left[\int y\hat{f}_k(y)dy\right]$$

$$= \int y\left[\frac{1}{nh_n}\sum_{t=1}^{n} - \frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}K\left(\frac{X_t - y}{sh_n}\right)\right]dy$$

$$= \frac{1}{n}\sum_{t=1}^{n} - \frac{1}{c_{k,0}}\sum_{|s|=1}^{k} c_{k,s}\int (X_t + sh_n\psi)K(\psi)\,d\psi = \frac{1}{n}\sum_{t=1}^{n} X_t.$$

Note that $E\left[\hat{f}_k(x) - f(x)\right] \le Ch_n^r\|f\|_{B_{\infty,q}^r}$ and $E\left[\hat{F}_k(x) - F(x)\right] \le Ch_n^r\|F\|_{B_{\infty,q}^r}$. Therefore $\hat{a}_k(x)$ is asymptotically unbiased.

Now I consider $Var(\hat{a}_k(y))$.

$$Var(\hat{a}_k(y)) = Var\left(\hat{\mu} - y + 2\hat{F}_k(y) - 2y\int_{-\infty}^{y} x\hat{f}_k(x)dx\right)$$

$$= Var\left(\hat{\mu} + 2\hat{F}_k(y) - 2y\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right)$$

$$= Var(\hat{\mu}) + 4Var\left(\hat{F}_k(y)\right) + 4y^2Var\left(\int_{-\infty}^{y} x\hat{f}_k(x)dx\right) + 4Cov(\hat{\mu}, \hat{F}_k(y)) - 4yCOV\left(\hat{\mu}, \int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right)$$
$$- 8yCov\left(\hat{F}_k(y), 2y\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right)$$

where $\hat{\mu} = \int x\hat{f}_k(x)dx$.

$$Var(\hat{\mu}) = Var\left(\frac{1}{n}\sum_{t=1}^{n} X_t\right) = \frac{1}{n^2}\sum_{t=1}^{n} Var(X_t) = \frac{\sigma^2}{n}$$

$$Var(\hat{F}_k(x)) = E\left[\hat{F}_k^2(x)\right] - \left(E\left[\hat{F}_k(x)\right]\right)^2$$

$$= \int \left[\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{x-\phi}{h_n}} M_k(\psi)\,d\psi\right]^2 f(\phi)\,d\phi - \left[\int \frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{x-\phi}{h_n}} M_k(\psi)\,d\psi f(\phi)\,d\phi\right]^2$$

$$= \frac{1}{n}\int \mathcal{M}_k^2\left(\frac{x-\phi}{h_n}\right) f(\phi)\,d\phi - \frac{1}{n}\left[\int \mathcal{M}_k\left(\frac{x-\phi}{h_n}\right) f(\phi)\,d\phi\right]^2$$

$$= \frac{1}{n}\int \mathcal{M}_k^2(\psi) f(x-h_n\psi)h_n\,d\psi - \frac{1}{n}\left[\int \mathcal{M}_k(\psi) f(x-h_n\psi)h_n\psi\right]^2$$

$$= -\frac{1}{n}\int \mathcal{M}_k^2(\psi)\,dF(x-h_n\psi) - \frac{1}{n}\left[-\int \mathcal{M}_k(\psi)\,dF(x-h_n\psi)\right]^2$$

$$= \frac{1}{n}\left\{2\int \mathcal{M}_k(\psi)M_k(\psi)F(x-h_n\psi)\,d\psi - \left[\int \mathcal{M}_k(\psi)F(x-h_n\psi)\,d\psi\right]^2\right\} = O(n^{-1})$$

$$Var\left(\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right)$$

$$= E\left[\left\{\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right\}^2\right] - \left(E\left[\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right]\right)^2$$

$$= E\left[\left\{\int_{-\infty}^{y} x\frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right)dx\right\}^2\right] - \left(E\left[\int_{-\infty}^{y} x\left(\frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right)\right)dx\right]\right)^2$$

$$= \frac{1}{n^2h_n^2}E\left[\left\{\int_{-\infty}^{y}\sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right)x\,dx\right\}^2\right] - \frac{1}{n^2h_n^2}\left(E\left[\left\{\int_{-\infty}^{y}\sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right)x\,dx\right\}\right]\right)^2$$

$$= \frac{n}{n^2h_n^2}\int\left[\int_{-\infty}^{y} M_k\left(\frac{x-X_1}{h_n}\right)x\,dx\right]^2 f(X_1)\,dX_1 + \frac{n(n-1)}{n^2h_n^2}\left[\int\int_{-\infty}^{y} M_k\left(\frac{x-X_1}{h_n}\right)x\,dx f(X_1)\,dX_1\right]^2$$
$$- \frac{1}{h_n^2}\left[\int\int_{-\infty}^{y} M_k\left(\frac{x-X_1}{h_n}\right)x\,dx f(X_1)\,dX_1\right]^2$$

$$= \frac{1}{n}\int\left\{\int_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi)(X_1+h_n\psi)\,d\psi\right\}^2 f(X_1)\,dX_1$$
$$- \frac{1}{n}\left(\int\left\{\int_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi)(X_1+h_n\psi)\,d\psi\right\} f(X_1)\,dX_1\right)^2 = O(n^{-1})$$

since $\int|K(\psi)|<\infty$, $\int|K(\psi)\psi|d\psi<\infty$ and $\int|\psi|f(\psi)d\psi<\infty$.

Let $\mathcal{M}_k(\psi)=\int_{-\infty}^{\psi}K(x)\,dx$.

$$Cov(\hat{\mu},\ \hat{F}_k(y))$$

$$= E[\hat{\mu}\hat{F}_k(y)] - E[\hat{\mu}]E[\hat{F}_k(y)]$$

$$= E\left[\left(\frac{1}{n}\sum_{t=1}^{n} X_t\right)\left(\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{y-X_t}{h_n}} M_k(\psi)d\psi\right)\right] - E\left[\frac{1}{n}\sum_{t=1}^{n} X_t\right] E\left[\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{y-X_t}{h_n}} M_k(\psi)\,d\psi\right]$$

$$= E\left[\left(\frac{1}{n}\sum_{t=1}^{n} X_t\right)\left(\frac{1}{n}\sum_{t=1}^{n} \mathcal{M}_k\left(\frac{y-X_t}{h_n}\right)\right)\right] - E\left[\left(\frac{1}{n}\sum_{t=1}^{n} X_t\right)\right] E\left[\frac{1}{n}\sum_{t=1}^{n} \mathcal{M}_k\left(\frac{y-X_t}{h_n}\right)\right]$$

$$= \frac{1}{n^2} E\left[\sum_{t=1}^{n} X_t \mathcal{M}_k\left(\frac{y-X_t}{h_n}\right)\right] + \frac{n(n-1)}{n^2} E[X_1] E\left[\mathcal{M}_k\left(\frac{y-X_t}{h_n}\right)\right] - E[X_1] E\left[\mathcal{M}_k\left(\frac{y-X_1}{h_n}\right)\right]$$

$$= \frac{1}{n} E\left[X_1 \mathcal{M}_k\left(\frac{y-X_1}{h_n}\right)\right] - \frac{1}{n} E[X_1] E\left[\mathcal{M}_k\left(\frac{y-X_1}{h_n}\right)\right]$$

$$= \frac{1}{n}\int X_1 \mathcal{M}_k\left(\frac{y-X_1}{h_n}\right) f(X_1)\,dX_1 - \frac{1}{n} E[X_1]\int \mathcal{M}_k\left(\frac{y-X_1}{h_n}\right) f(X_1)\,dX_1 = O(n^{-1})$$

The above equation holds by the inequality $|\mathcal{M}_k(\psi)| < \infty$.

$$|\mathcal{M}_k(\psi)| = \left|\int_{-\infty}^{\psi} M_k(x)\,dx\right| = \left|-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\int_{-\infty}^{\psi} K\left(\frac{\phi}{s}\right)d\phi\right| \le \left|-\frac{1}{c_{k,0}}\right|\left|\sum_{|s|=1}^{k} c_{k,s}\right|\int |K(\phi)|\,d\phi < \infty$$

$$Cov\left(\hat{\mu},\ \int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right)$$

$$= E\left[\hat{\mu}\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right] - E[\hat{\mu}] E\left[\int_{-\infty}^{y} x\hat{f}_k(x)\,dx\right]$$

$$= E\left[\frac{1}{n}\sum_{t=1}^{n} X_t \int_{-\infty}^{y}\frac{x}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right)dx\right] - E\left[\frac{1}{n}\sum_{t=1}^{n} X_t\right] E\left[\int_{-\infty}^{y}\frac{x}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right)dx\right]$$

$$\begin{aligned}= \frac{1}{n^2 h_n} E\left[\sum_{t=1}^{n} X_t \int_{-\infty}^{y} x M_k\left(\frac{x-X_t}{h_n}\right)dx\right] &+ \frac{n(n-1)}{n^2 h_n} E[X_1] E\left[\int_{-\infty}^{y} x\, M_k\left(\frac{x-X_1}{h_n}\right)dx\right] \\ &- E[X_1] E\left[\int_{-\infty}^{y}\frac{x}{h_n} M_k\left(\frac{x-X_1}{h_n}\right)dx\right]\end{aligned}$$

$$= \frac{1}{nh_n} E\left[X_1 \int_{-\infty}^{y} x M_k\left(\frac{x-X_1}{h_n}\right)dx\right] - \frac{1}{nh_n} E[X_1] E\left[\int_{-\infty}^{y} x M_k\left(\frac{x-X_1}{h_n}\right)dx\right]$$

$$= \frac{1}{n}\int X_1 \int_{-\infty}^{\frac{y-X_1}{h_n}} (X_1 + h_n\psi) M_k(\psi) f(X_1)\,d\psi dX_1 - \frac{1}{n} E[X_1]\int_{-\infty}^{\frac{y-X_1}{h_n}} (X_1 + h_n\psi) M_k(\psi) f(X_1)\,d\psi dX_1$$

$$= \frac{1}{n}\iint_{-\infty}^{\frac{y-X_1}{h_n}} (X_1^2 + X_1 h_n\psi) M_k(\psi)\, f(X_1)\,d\psi dX_1 - \frac{1}{n} E[X_1]\iint_{-\infty}^{\frac{y-X_1}{h_n}} (X_1 + h_n\psi) M_k(\psi) f(X_1)\,d\psi dX_1$$

$$= \frac{1}{n}\int X_1^2 \int_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi) f(X_1) d\psi dX_1$$

$$+ \frac{1}{n}\int X_1 \int_{-\infty}^{\frac{y-X_1}{h_n}} h_n \psi M_k(\psi) f(X_1) d\psi dX_1$$

$$- \frac{1}{n} E[X_1] \int X_1 \int_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi) f(X_1) d\psi dX_1 - \frac{1}{n} E[X_1] \iint_{-\infty}^{\frac{y-X_1}{h_n}} h_n \psi M_k(\psi) f(X_1) d\psi dX_1$$

$$= O(n^{-1})$$

$$Cov\left(\hat{F}_k(y), \int_{-\infty}^{y} x \hat{f}_k(x) dx\right)$$

$$= Cov\left[\int_{-\infty}^{y} \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right) dx, \quad \int_{-\infty}^{y} \frac{1}{nh_n} x \sum_{t=1}^{n} M_k\left(\frac{x-X_t}{h_n}\right) dx\right]$$

$$= E\left[\frac{1}{n^2 h_n^2} \sum_{t=1}^{n} \int_{-\infty}^{y} M_k\left(\frac{x-X_t}{h_n}\right) dx \int_{-\infty}^{y} x M_k\left(\frac{x-X_t}{h_n}\right) dx\right]$$

$$- \frac{1}{nh_n^2} E\left[\int_{-\infty}^{y} M_k\left(\frac{x-X_1}{h_n}\right) dx\right] E\left[\int_{-\infty}^{y} x M_k\left(\frac{x-X_1}{h_n}\right) dx\right]$$

$$= \frac{1}{n}\iint_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi) d\psi \int_{-\infty}^{\frac{y-X_1}{h_n}} (X_1 + h_n\psi) M_k(\psi) d\psi f(X_1) dX_1$$

$$- \frac{1}{n}\iint_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi) d\psi f(X_1) dX_1 \iint_{-\infty}^{\frac{y-X_1}{h_n}} (X_1 + h_n\psi) M_k(\psi) d\psi f(X_1) dX_1$$

$$= O(n^{-1})$$

provided by $\left|\int_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi) d\psi\right| \leq \int |M_k(\psi) d\psi| d\psi < \infty$ and $\int |X_1| f(X_1) dX_1 < \infty.$

Therefore, I conclude following:

$Var(\hat{a}_k(y)) = O(n^{-1})$

Proof of Theorem 3

$$[\hat{P}_\alpha(\hat{F}_k) - P_\alpha(F)]$$
$$= \int \hat{f}_k(y)^\alpha \hat{a}_k(y) d\hat{F}_k(y) - \int f(y)^\alpha a(y) dF(y)$$
$$= \int \hat{p}_\alpha(y) d\hat{F}(y) - \int p_\alpha(y) dF(y)$$
$$= \int [\hat{p}_\alpha(y) - p_\alpha(y)] dF(y) + \int p_\alpha(y) d[\hat{F}_k(y) - F(y)] + \int [\hat{p}_\alpha(y) - p(y)] d[\hat{F}_k - F](y)$$

where $\hat{p}_\alpha \equiv \hat{f}_k^\alpha(y)\hat{a}_k(y)$ and $p_\alpha \equiv f(y)^\alpha a(y)$.

$$\hat{p}_\alpha(y) - p_\alpha(y)$$
$$= \hat{f}_\alpha^\alpha(y)\hat{a}_k(y) - f(y)^\alpha a(y)$$
$$= [\hat{f}_k(y)^\alpha - f(y)^\alpha][\hat{a}_k(y) - a(y)] + [\hat{f}_k(y)^\alpha - f(y)^\alpha] a(y) + f(y)^\alpha [\hat{a}_k(y) - a(y)]$$
$$= \alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)] a(y) + \alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)][\hat{a}_k(y) - a(y)] + f(y)^\alpha [\hat{a}_k(y) - a(y)]$$

According to the mean value theorem, I have $\hat{f}_k(y)^\alpha - f(y) = \alpha f^*(y)^{\alpha-1}\left(\hat{f}_k(y) - f(y)\right)$ where $f^*(y)$ lies in between f and $\hat{f}_k$ for all $y \in \mathbb{R}$.

$$E[\hat{P}_k(\hat{F}_k) - P_\alpha(F)]$$
$$= E\left[\int [\hat{p}_\alpha(y) - p_\alpha(y)] dF(y)\right] + E\left[\int p_\alpha(y) d[\hat{F}_k(y) - F(y)]\right] + E\left[\int [\hat{p}_\alpha(y) - p_\alpha(y)] d[\hat{F}_k - F](y)\right]$$

$$E\left[\int p_\alpha(y) d[\hat{F}_k(y) - F(y)]\right] = \int E\left[p_\alpha(y)[\hat{f}_k(y) - f(y)]\right]$$
$$= \int p_\alpha(y) E[\hat{f}_k(y) - f(y)] dy \le O(h_n^r) E[p_\alpha(y)]$$

Note that

$$E\left[\left(\hat{f}_k(y) - f(y)\right)^2\right] = Bias\left(\hat{f}_k(y)\right)^2 + Var\left(\hat{f}_k(y)\right) = O(h_n^{2r}) + O((nh_n)^{-1})$$

and

$E\left[\left(\hat{a}_k(y)-a(y)\right)^2\right]=Bias\left(\hat{a}_k(y)\right)^2+Var\left(\hat{a}_k(y)\right)$. Hence,

$|E[\int[\hat{p}_\alpha(y)-p_\alpha]dF(y)]|\to 0$ as $n\to\infty$.

Similarly, I have as $n\to\infty$

$$E\left[\int[\hat{p}_\alpha(y)-p_\alpha]d[\hat{F}_k-F](y)\right]$$

$$=\int\left\{E\left[\left(\hat{p}_\alpha(y)-p_\alpha(y)\right)^2\right]\right\}^{1/2}\left\{E\left[\left(\hat{f}_k(y)-f(y)\right)^2\right]\right\}^{1/2}dy\longrightarrow 0$$

Now consider the part (b) then I need to show that $Var(\hat{P}_\alpha(\hat{F}_k)\longrightarrow 0$ as $n\to\infty$.

$$Var\left(\hat{P}_\alpha(\hat{F}_k)\right)$$

$$=Var\left(\int\hat{a}_k(y)\hat{f}_k(y)^{1+\alpha}dy\right)$$

$$=Var\left(\int\hat{a}_k(y)[f(y)^{1+\alpha}+(1+\alpha)f^*(y)^\alpha(\hat{f}_k(y)-f(y))]dy\right)$$

$$=Var\left(\int\hat{a}_k(y)f(y)^{1+\alpha}dy+(1+\alpha)\int\hat{a}_k(y)f^*(y)^\alpha\hat{f}_k(y)dy-(1+\alpha)\int\hat{a}_k(y)f^*(y)^\alpha f(y)dy\right)$$

$$\int\hat{a}_k(y)f(y)^{1+\alpha}dy$$

$$=\int\left[\hat{\mu}-y+2\hat{F}_k(y)-2y\int_{-\infty}^{y}x\hat{f}_k(x)dx\right]f(y)^{1+\alpha}dy$$

$$=\frac{1}{n}\sum_{t=1}^{n}X_t\int f(y)^{1+\alpha}dy-\int yf(y)^{1+\alpha}dy+2\int\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)\,d\psi f(y)^{1+\alpha}dy$$
$$-2\int y\int_{-\infty}^{y}x\left[\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{x-X_t}{h_n}\right)\right]dxf(y)^{1+\alpha}dy$$

$$=\frac{1}{n}\sum_{t=1}^{n}X_t\int f(y)^{1+\alpha}dy-\int yf(y)^{1+\alpha}dy+\frac{2}{n}\sum_{t=1}^{n}\iint_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)\,d\psi f(y)^{1+\alpha}dy$$

$$-\frac{2}{n}\sum_{t=1}^{n}\int y\int_{-\infty}^{\frac{y-X_t}{h_n}}(X_t+h_n\psi)M_k(\psi)d\psi f(y)^{1+\alpha}dy$$

$$=\frac{1}{n}\sum_{t=1}^{n}X_t\int f(y)^{1+\alpha}dy-\int yf(y)^{1+\alpha}dy+\frac{2}{n}\sum_{t=1}^{n}\iint_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)\,d\psi f(y)^{1+\alpha}dy$$

$$-\frac{2}{n}\sum_{t=1}^{n}X_t\int y\int_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)d\psi f(y)^{1+\alpha}dy-\frac{2}{n}\sum_{t=1}^{n}h_n\int y\int_{-\infty}^{\frac{y-X_t}{h_n}}\psi M_k(\psi)\,d\psi f(y)^{1+\alpha}dy$$

$$(1+\alpha)\int \hat{a}_k(y)f^*(y)^{\alpha}\hat{f}_k(y)dy$$

$$=(1+\alpha)\int\left[\hat{\mu}-y+2\hat{F}_k(y)-2y\int_{-\infty}^{y}x\hat{f}_k(x)dx\right]f^*(y)^{\alpha}\hat{f}_k(y)dy$$

$$=(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}X_t\int f^*(y)^{\alpha}\left[\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{y-X_t}{h_n}\right)\right]dy-(1+\alpha)\int yf^*(y)^{\alpha}\left[\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{y-X_t}{h_n}\right)dy\right]$$

$$+2(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\iint_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)\,d\psi f^*(y)^{\alpha}\left[\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{y-X_t}{h_n}\right)\right]dy$$

$$-2(1+\alpha)\int y\int_{-\infty}^{y}x\left[\frac{1}{nh_n}M_k\left(\frac{x-X_t}{h_n}\right)\right]dxf^*(y)^{\alpha}\left[\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{y-X_t}{h_n}\right)\right]dy$$

$$=(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}X_t\int f^*(X_t+h_n\psi)^{\alpha}M_k(\psi)d\psi-(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\int(X_t+h_n\psi)f^*(X_t+h_n\psi)^{\alpha}M_k(\psi)d\psi$$

$$+2(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\iint_{-\infty}^{\theta}M_k(\psi)d\psi f^*(X_t+h_n\theta)^{\alpha}M_k(\theta)\,d\theta$$

$$-2(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\int(X_t+h_n\theta)\left[\int_{-\infty}^{\theta}(X_t+h_n\psi)M_k(\psi)\,d\psi\right]f^*(X_t+h_n\theta)^{\alpha}M_k(\theta)d\theta$$

$$-(1+\alpha)\int \hat{a}_k(y)f^*(y)^{\alpha}f(y)dy$$

$$=-(1+\alpha)\int\left[\hat{\mu}-y+2\hat{F}_k(y)-2y\int_{-\infty}^{y}x\hat{f}_k(x)dx\right]f^*(y)^{\alpha}f(y)dy$$

$$=-(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}X_t\int f^*(y)^{\alpha}f(y)dy+(1+\alpha)\int yf^*(y)^{\alpha}f(y)\,dy$$

$$-2(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\iint_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)\,d\psi f^*(y)^{\alpha}f(y)\,dy$$

$$+2(1+\alpha)\int y\int_{-\infty}^{y}x\left[\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{x-X_t}{h_n}\right)\right]dxf^*(y)^{\alpha}f(y)\,dy$$

$$=-(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}X_t\int f^*(y)^{\alpha}f(y)dy+(1+\alpha)\int yf^*(y)^{\alpha}f(y)\,dy$$

$$-2(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\iint_{-\infty}^{\theta} M_k(\psi)\,d\psi\, f^*(X_t+h_n\theta)^{\alpha} f(X_t+h_n\theta)\,h_n\,d\theta$$

$$+2(1+\alpha)\frac{1}{n}\sum_{t=1}^{n}\int (X_t+h_n\theta)\int_{-\infty}^{\theta}(X_t+h_n\psi)M_k(\psi)\,d\psi\, f^*(X_t+h_n\theta)^{\alpha} f(X_t+h_n\theta)\,h_n\,d\theta$$

$$nVar\big(\hat{P}_{\alpha}(\hat{F}_k)\big)$$

$$\longrightarrow Var\Big(X_1\int f(y)^{1+\alpha}dy-\int yf(y)^{1+\alpha}dy+2\int f(y)^{1+\alpha}dy-2X_1\int yf(y)^{1+\alpha}dy$$

$$+2(1+\alpha)f(X_1)^{\alpha}\iint_{-\infty}^{\theta} M_k(\psi)\,d\psi M_k(\theta)\,d\theta-2(1+\alpha)X_1^2 f(X_1)^{\alpha}\iint_{-\infty}^{\theta} M_k\,d\psi M_k(\theta)\,d\theta$$

$$-(1+\alpha)X_1\int f(y)^{\alpha}f(y)\,dy+(1+\alpha)\int yf(y)^{1+\alpha}dy\Big)$$

$$=Var\Big(-\alpha X_1\int f(y)^{1+\alpha}dy-2X_1\int yf(y)^{1+\alpha}dy+2(1+\alpha)f(X_1)^{\alpha}(1-X_1^2)\iint_{-\infty}^{\theta} M_k(\psi)\,d\psi M_k(\theta)\,d\theta\Big)$$

Proof. Theorem 4

$$\hat{P}_k(\hat{F}_k)$$
$$= \int \hat{a}_k(y)\hat{f}_k(y)^{1+\alpha}dy$$
$$= \int \hat{a}_k(y)f(y)^{1+\alpha}dy + (1+\alpha)\int \hat{a}_k(y)f^*(y)^{\alpha}\hat{f}_k(y)dy - (1+\alpha)\int \hat{a}_k(y)f^*(y)^{\alpha}f(y)dy$$

Let $\hat{P}_k(\hat{F}_k) \equiv \Sigma_{t=1}^{n} Z_{nt}$

$$Z_{nt}$$
$$= \frac{1}{n}\Big[X_t \int f(y)^{1+\alpha}dy - \int yf(y)^{1+\alpha}dy + 2\iint_{-\infty}^{\theta} M_k(\psi)\,d\psi f(X_t + h_n\psi)M_k(\psi)\,d\psi f(X_t + h_n\theta)^{1+\alpha}h_n d\theta$$
$$-2\int (X_t + h_n\theta)\int_{-\infty}^{\theta}(X_t + h_n\psi)M_k(\psi)\,d\psi f(X_t + h_n\theta)^{1+\alpha}h_n d\theta$$
$$+(1+\alpha)X_t\left(\frac{1}{n}\sum_{t=1}^{n}\int f^*(X_t + h_n\psi)^{\alpha}M_k(\psi)\,d\psi\right)$$
$$-(1+\alpha)\int (X_t + h_n\psi)f^*(X_t + h_n\psi)^{\alpha}M_k(\psi)\,d\psi + 2(1+\alpha)\iint_{-\infty}^{\theta} M_k(\psi)d\psi f^*(X_t + h_n\theta)^{\alpha}M_k(\theta)\,d\theta$$
$$-2(1+\alpha)\int (X_t + h_n\theta)\int_{-\infty}^{\theta}(X_t + h_n\psi)M_k(\psi)\,d\psi f^*(X_t + h_n\theta)^{\alpha}M_k(\theta)d\theta - (1+\alpha)X_t\int f^*(y)^{\alpha}f(y)\,dy$$
$$+(1+\alpha)\int yf^*(y)^{\alpha}f(y)\,dy - 2(1+\alpha)\iint_{-\infty}^{\theta} M_k(\psi)\,d\psi f^*(X_t + h_n\theta)^{\alpha}f(X_t + h_n\theta)\,h_n d\theta$$
$$+2(1+\alpha)\int (X_t + h_n\theta)\int_{-\infty}^{\theta} M_k(\psi)\,d\psi f^*(X_t + h_n\theta)^{\alpha}f(X_t + h_n\theta)h_n d\theta\Big]$$

Hence, I have $E[Z_{nt}] = O(n^{-1})$.

$$\sum_{t=1}^{n} E\left[\left(Z_{nt} - E(Z_{nt})\right)^2\right] = S_n^2 = Var\left(\hat{P}_\alpha(\hat{F}_k)\right)$$

Note that

$$\frac{\hat{P}_\alpha(\hat{F}_k) - E[\hat{P}_k(\hat{F}_k)]}{\sqrt{Var\left(\hat{P}_k(\hat{F}_k)\right)}} = \sum_{t=1}^{n}\left[\frac{Z_{nt} - E[Z_{nt}]}{S_n}\right] = \sum_{t=1}^{n} X_{nt}.$$

Hence, $E[X_{nt}] = 0$, $E[X_{nt}^2] = \frac{1}{S_n^2}E\left[\left(Z_{nt} - E(Z_{nt})\right)^2\right]$ and $\Sigma_{t=1}^{n} E[X_{nt}^2] = 1$.

To satisfy Liapounov's CLT, I need to know that $\lim_{n\to\infty}\sum_{t=1}^{n} E|X_{nt}|^{2+\delta}=0$. Let $Z_{nt}=\frac{1}{n}B_t$.

$$\sum_{t=1}^{n} E|X_{nt}|^{2+\delta}=\sum_{t=1}^{n} E\left[\left|\frac{Z_{nt}-E[Z_{nt}]}{S_n}\right|^{2+\delta}\right]$$

$$=\sum_{t=1}^{n} Var(S_n)^{-1-\delta/2} E\left[|Z_{nt}-E[Z_{nt}]|^{2+\delta}\right]$$

$$\leq \sum_{t=1}^{n} Var(S_n)^{-1-\delta/2} 2^{1+\delta} E\left[|Z_{nt}|^{2+\delta}+o(1)\right] \qquad \text{by using } C_r \text{ inequality}$$

$$\leq nVar(S_n)^{-1-\delta/2} 2^{1+\delta} n^{-2-\delta}+E\left[|B_t|^{2+\delta}+o(1)\right]$$

$$=Var(S_n)^{-1-\delta/2} 2^{1+\delta} n^{-1-\delta} E\left[|B_t|^{2+\delta}+o(1)\right]$$

$$=\frac{1}{\left(nVar(S_n)\right)^{1+\delta/2}} 2^{1+\delta} \frac{1}{n^{\delta/2}} E\left[|B_t|^{2+\delta}+o(1)\right]$$

From the previous theorem, I know that $nVar(S_n) \longrightarrow var$ as $n\to\infty$. Then,

$$\lim_{n\to\infty}\sum_{t=1}^{n} E|X_{nt}|^{2+\delta}=0.$$

Hence, I have

$$\sqrt{n}\left[\hat{P}_\alpha(\hat{F}_k)-E[\hat{P}_\alpha(\hat{F}_k)]\right] \xrightarrow{d} N(0,\ var).$$

양극화 지수 추정: 비모수 커널 추정법을 기반으로

이나경*

논문요약

본 연구에서는 Duclos 등이 제시한 양극화 지수(DER Index)의 새로운 추정방법을 제시하고 있다. 기존 DER Index에 대한 추정치는 경험적 분포함수(empirical distribution function)와 Rosenblatt-Parzen 커널 밀도 추정함수(kernel density estimator)에 기초하고 있다. 그러나 경험적 분포함수를 이용한 기존의 DEX Index 추정은 평탄성(smoothness)의 부재라는 한계가 있다. 따라서 본고에서는 기존의 DER Index 추정치의 한계를 해결하기 위해 새로운 추정치를 제안한다. 이 추정치는 Mynbaev and Martins-Filho가 제시한 새로운 형태의 커널 밀도 추정치(kernel density estimator)를 기반으로 도출되어 기존 추정치의 한계를 극복하였다. 또한 이 추정치는 점근적 일치(consistent), $\sqrt{n}$ 속도로 점근정규성(asymptotic normality)을 확립하고 있음이 증명되었다. 몬테카를로 시뮬레이션 연구결과, 본고에서 제시한 추정치가 Bias와 평균제곱오차 측면에서 기존의 DER Index 추정치에 비해 우수한 것으로 입증되었다.

주제어: 양극화 측정값, DER 지수, 비모수 밀도 함수

투고일: 2021.05.14. 심사일: 2021.06.21. 게재확정일: 2021.07.04.

* 서울여자대학교 경제학과, 조교수.