What Aspect of the Income Distribution Matters in the Crime Model:

Bipolarization Index and Its Inferences

Yoonseok Lee

Department of Economics, University of Michigan

yoolee@umich.edu

Donggyun Shin

Department of Economics, Kyung Hee University

dgshin@khu.ac.kr

First draft: *June 30, 2007*

This version: *February 12, 2008*

[*Very Preliminary; comments are welcome*]

Abstract

The main purpose of this paper is to investigate what aspect of the income distribution matters in explaining the crime rate. It is commonly believed that crime rates are positively associated with the degree of income inequality. Unlike the conventional wisdom, we emphasize the bipolarization nature of an income distribution as a crime determinant. It is hypothesized that, as the income distribution becomes more bipolarized, mobility between different income groups is reduced, which lowers both the expected future income of the poor and their incentive to supply labor in the legal labor market. As a result, with other things being constant including the degree of income inequality, crime rates are expected to increase as the income distribution is more bipolarized.

One important aspect of such hypothesis is to assume asymmetric degrees of crime incentive between the poor and the rich. For the purpose of testing the hypothesis, the bipolarization measure of Esteban, Gradin, and Ray (1999) is generalized by allowing for asymmetric degrees of alienation feeling (or social antagonism) between the low- and the high-income classes. Fixed effects GMM estimates based on the U.S. state level data from 1991 through 2005 suggest that, when both the Gini and the bipolarization indices are included in the regression model, bipolarization significantly raises both the property and the violent crime rates, whereas the Gini index is not related with any type of crime. More importantly, consistent with our theoretical prediction regarding asymmetric alienation feeling between the rich and the poor, the explanatory power of our generalized bipolarization index increases as a heavier weight is placed on the alienation feeling of the lower income class.

1 Background

Much effort has been made to explain theoretically and empirically the linkage between inequality and crime. As well summarized by Kelly (2000), Fajnzylber, Lederman, and Loayza (2002), and Demombynes and Ozler (2005) among others, the relationship has been the subject of various economic and sociological theories on crime. While the Beckerian economic models of crime−the more unequal an income distribution is, the greater the gap between benefits and costs of crime, and thus the higher the (property) crime rate is−are largely based on the differential returns from legal and illegal activities, economic and sociological theories vary depending on which aspect of social phenomena is operative in their theories. According to sociological theories, inequality generally undermines the society's ability to control its members, reduces demand for public demand for safety, or reduces social mobility, which all work in the direction of raising the crime rate. As well argued by Fajnzylber, Lederman, and Loayza (2002, p.2), however, it is difficult to empirically identify whether the positive crime-inequality relation, if any, results from economic incentives-disincentives of crime or from social strain or disorganization.

Turning to empirical evidence, most cross-sectional comparisons across states and cities in the United States or across countries conclude that inequality leads to property and/or violent crime.[1] Findings in these cross-sectional studies, however, may be subject to an omitted variable bias problem, as they do not control for unobserved fixed effects that are specific to the cross-sectional unit and possibly correlated with the unit's inequality. Even for panel data models with fixed effects, however, evidence is still weak and mixed. For example, Lee (1993, cited in Freeman, 1996) regresses changes in crime among metropolitan area between 1970 and 1980 against changes in inequality and finds insignificant coefficient estimates. Doyle, Ahmed, and Horn's (1999) first-differenced model also produces insignificant coefficients of the Gini index. On the contrary, on the basis of international panel data for 39 countries, Fajnzylber, Lederman, and Loayza (2002) report significantly positive co-

[1] See Demombynes and Ozler (2005) for the most comprehensive up-to-date survey of these cross-sectional studies.

efficients for both homicide rates and robbery rates even after country-specific fixed-effects are controlled for.

How to represent income inequality is one of the central issues in this literature. Various inequality measures have been tested as potential crime determinants. For example, some researchers use the Gini coefficient (e.g., Ehrlich, 1973; Blau and Blau, 1982; Fajnzylber, Lederman, Loayza, 2002), others use the proportion of the population below a certain percentage of the median income (Nilsson, 2004; Bourguignon, Nunez, and Sanchez, 2003),[2] while another exploits the mean log deviation as a special case of Generalized Entropy measure (Demombynes and Ozler, 2005). It is possible that, as formally addressed by Bourguignon, Nunez, and Sanchez (2003), results may be different depending on which part of the income distribution is used to measure the degree of income inequality. Unlike these studies, this paper focuses on the bipolarization aspect of the income distribution as a determinant of individuals' crime. While inequality represents dispersion around the global mean, bipolarization emphasizes disappearing middle class and clustering around two local means, which has a strong implication on income mobility.

Why should bipolarization matter in the crime model? What are economic and social consequences of bipolarization? More specifically, what additional role other than inequality does bipolarization play in explaining crime behaviors? First, as already stressed by Foster and Wolfson (1992), Wolfson (1994), Wolfson (1997) and Esteban and Ray (1994), polarization and inequality are conceptually different. Two distributions with the same degree of inequality may reveal different levels of bipolarization. Bipolarization emphasizes, in addition to the distance between different income groups, within-group clustering. Second, polarization has important implications on political cohesion and democratic decision making. The fact that the distribution of individuals' attributes in a society is bipolarized implies that a social consensus on the same subject is costly. Indeed, the polarization indices developed by Esteban and Ray (1994) and Esteban, Gradin, and Ray (1999) are intended to reflect the level of social conflict or social unrest in general. Esteban and Ray (1999) try

[2] Nilsson(2004) finds that the proportion of the population with an income below 10 percent of median income matters most in explaining the incidence of property crime.

to identify the type of distributions under which social conflict is most likely. Truly, the conventional inequality may not be appropriate in explaining this type of 'collective crime.' The very differences manifested in great inequalities tend to deprive the lower strata of the strength and resources to organize successful collective action (Blau and Blau, 1982, p.119).

Little attempt, however, has been made to empirically support the polarization-conflict relationship, presumably because of the data problem. Not enough observations exist that reflect conflictions between groups such as events of riots or revolts, although history has witnessed them occasionally. This paper derives a testable hypothesis regarding economic and social impacts of bipolarization. Unlike Esteban and Ray (1999), we attempt to explain how bipolarization affects 'individual crime behavior.' As will be discussed subsequently, for a given level of overall inequality, a more bipolarized society has less mobility between the two income groups, which works in the direction of increasing crime incentives of low income earners.

This paper organized as follows. Section 2 describes our simple hypothesis regarding the crime-polarization relationship. Section 3 develop a bipolarization index to better represent the total effective antagonism that individuals have in a geographical region. An empirical crime model along with the estimation strategy, data, and empirical findings are presented in Section 4. Section 5 concludes the paper by summarizing the main results. Technical results are summarized in Appendix.

2 A Hypothesis

When a person decides, at a point in time, whether to supply his labor in the legal or illegal labor market, the person considers his relative position not just in the current distribution but also in the distribution of expected life time income. While the former is related with the conventional static inequality measure such as the Gini coefficient, the latter is determined by the person's entire life time income stream, which in turn depends on, among others, the person's mobility to the other income group than the one he belongs to. Given this perspective, even a low income earner at a certain point in time would not feel a great crime incentive if he had better prospects in the future. The greater the mobility, the higher the person's expected lifetime income, and the weaker crime incentive he has. Of course, from a high income earner's point of view, mobility to the lower income group would reduce her life time income, and incentive to supply labor in the legal market, although this increased crime incentive of the high income earner is presumably smaller in magnitude than the reduced crime incentive of the low income earner.

The mobility implied by an income distribution is better summarized by polarization rather than inequality. In particular, the concept and index developed by Esteban, Gradin, and Ray (henceforth EGR, 1999) have strong implications on income mobility between groups, although this aspect is not explicitly emphasized in their work. According to EGR, an income distribution is more polarized when either the between group distance is greater or within group clustering becomes stronger. We postulate that mobility decreases in between group distance and within group clustering. From the viewpoint of a lower income earner, when mobility were to increase or the distribution were less bipolarized, expected future income would become higher, which would increase the marginal cost of crime, even though the degree of income inequality remained the same. From the high income earner's point of view, reduced mobility or more bipolarization would increase her expected life time income, which would *ceteris paribus* increase the marginal cost of crime of high income earners. It is hypothesized that, other things being constant, crime rates increase in the degree of bipolarization.

Alternatively, the overall crime rate in a society may depend on overall level of unhappiness or antagonism contained in a society. The EGR index is basically designed to represent the sum of all effective antagonism in a society, which is determined not only by the feeling of alienation each individual has against those in the income group he does not belong to; but also by the feeling of identification the person has toward individuals in his group. The more alienated, or the more identified, the more polarized. In this context, an enhanced (reduced) mobility would make either the identification or the alienation weaker (stronger), which reduces (increases) the degree of polarization, which in turn lowers (raises) the overall level of effective antagonism.

In the following section, the EGR index is extended to better represent the overall level of effective antagonism in a society. Unlike EGR, we will allow asymmetric alienation feelings between different income classes: low income earners feel more alienated from high income earners than high income earners do from low income earners. In addition, we focus on how polarization affects an individual's crime behavior instead of a group's collective action.

3 Measuring Effective Antagonism

3.1 The Setup

We assume that a set of individual income data $\{y_i\}_{i=1}^n$ is a random sample from an underlying distribution $F(y)$, whose support is given by $[y_{\min}, y_{\max}]$ with $0 < y_{\min} < y_{\max} < \infty$. We consider K number of pre-specified and disjoint intervals $[a_{k-1}, a_k)$ for $k = 1, 2, \cdots, K$ and $2 \leq K \leq n$; we let each observation y_i fall in one of the K intervals. Without loss of generality, we let $y_{\min} = a_0 < a_1 < \cdots < a_{K-1} < a_K = y_{\max}$ and define the last interval $[a_{K-1}, a_K]$ to be closed. The number of intervals, K, is given and it is assumed to be fixed (i.e., not growing with n) and small (e.g., $K = 2$ in the context of bipolarization).

In each interval $A_k = [a_{k-1}, a_k)$, we define the population fraction π_k, the group mean μ_k and the group variance σ_k^2 as

$$
\begin{aligned}
\pi_k &= \mathbb{P}\{y \in A_k\} = \int_{a_{k-1}}^{a_k} dF(y), \\
\mu_k &= \mathbb{E}[y|y \in A_k] = \frac{1}{\pi_k} \int_{a_{k-1}}^{a_k} y dF(y) \quad \text{and} \\
\sigma_k^2 &= var[y|y \in A_k] = \frac{1}{\pi_k} \int_{a_{k-1}}^{a_k} (y - \mu_k)^2 dF(y),
\end{aligned}
$$

where we assume that $\pi_k > 0$ and $0 < \sigma_k^2 < \infty$ for all k. It follows that $\sum_{k=1}^K \pi_k = 1$

and the overall mean is defined as $\mu = \sum_{k=1}^{K} \pi_k \mu_k = \int_{y_{\min}}^{y_{\max}} y dF(y)$. Notice that the group means are in ascending order by construction, i.e., $\mu_k < \mu_j$ if $k < j$. Using a random (viz., *i.i.d.*) sample $\{y_i\}_{i=1}^{n}$, we estimate π_k, μ_k and σ_k^2 by the sample proportion p_k, the group sample mean $\overline{y}_k$ and the group sample variance s_k^2, respectively, as follows:

$$\widehat{\pi}_k \equiv p_k = n_k/n, \tag{1}$$

$$\widehat{\mu}_k \equiv \overline{y}_k = (1/n_k) \sum_{i=1}^{n} y_i \mathbb{I}\{y_i \in A_k\} \tag{2}$$

$$\widehat{\sigma}_k^2 \equiv s_k^2 = (1/(n_k - 1)) \sum_{i=1}^{n} (y_i - \overline{y}_k)^2 \mathbb{I}\{y_i \in A_k\}, \tag{3}$$

where n_k is the number of observations in the interval A_k and $\mathbb{I}\{\cdot\}$ is the binary indicator.

3.2 Polarization Index

We consider a polarization index $\mathcal{P}$ given by[3]

$$\mathcal{P}(\alpha, \theta) = \frac{1}{2\mu} \sum_{k=1}^{K} \sum_{j=1}^{K} \pi_k^{1+\alpha} \pi_j \rho_\theta (\mu_k - \mu_j), \tag{4}$$

where $\alpha \geq 0$ and $0 \leq \theta \leq 1/2$ are parameters chosen by a researcher[4] and $\rho_\theta(u) = 2u(\theta - \mathbb{I}\{u < 0\})$. Notice that if $\theta = 1/2$, then $\rho_\theta(u) = |u|$ and the polarization index $\mathcal{P}(\alpha, \theta)$ in (4) is the same as Esteban and Ray (1994). In addition, if $\alpha = 0$ then the index becomes the Gini index from grouped data, $G = (1/2\mu) \sum_{k=1}^{K} \sum_{j=1}^{K} \pi_k \pi_j |\mu_k - \mu_j|$.

The polarization index $\mathcal{P}(\alpha, \theta)$ is based on the idea of Esteban and Ray (1994), whose

[3]It should be noted that the polarization index $\mathcal{P}(\alpha, \theta)$ satisfies the axioms of Duclos, Esteban and Ray (2004). They show that the general form of index satisfying the axioms is given as

$$DER(\alpha) = \int \int T(\mathcal{I}(x), \mathcal{A}(x, y)) dF(x) dF(y) = C \int \int \mathcal{I}(x)^\alpha \mathcal{A}(x, y) dF(x) dF(y)$$

with a constant C and α. Therefore, in a discrete expression of the integral operations, $\mathcal{P}(\alpha, \theta)$ can be understood as $DER(\alpha)$ if the identity function $\mathcal{I}(\cdot)$ is defined as $\mathcal{I}(x) = \pi_k \mathbb{I}\{x \in A_k\}$ and the alienation function $\mathcal{A}(\cdot, \cdot)$ is defined as $\mathcal{A}(x, y) = \rho_\theta(\mu_k - \mu_j) \mathbb{I}\{x \in A_k\} \mathbb{I}\{y \in A_j\}$ given θ.

[4]More precisely, Esteban and Ray (1994) restrict $0 < \alpha \leq 1.6$ to satisfy several axioms.

index is given by

$$ER_K(\alpha) = (1/\mu) \sum_{k=1}^{K} \sum_{j=1}^{K} \pi_k^{1+\alpha} \pi_j \left| \mu_k - \mu_j \right|.$$

Similarly, $\mathcal{P}(\alpha, \theta)$ also combines the following two concepts: within-group *identity* and the between-group *alienation*. The group identity is formulated as π_k^{α}, which is the same form as $ER_K(\alpha)$. The group alienation, on the other hand, is formulated as

$$\rho_\theta\left(\mu_k - \mu_j\right) = 2\left(\mu_k - \mu_j\right)\left(\theta - \mathbb{I}\left\{\mu_k < \mu_j\right\}\right),$$

which is more general than $ER_K(\alpha)$. Note that we allow for *asymmetric* feeling of alienation, where the degree of the asymmetry is determined by the value θ. Specifically, since we let $0 \leq \theta \leq 1/2$, the lower income groups feel more alienation to the higher income groups than how the higher income groups feel to the lower. The asymmetry gets severer as θ goes to zero and it can be understood as antagonism of the low income group to the high income group. As an extreme case, if $\theta = 0$ then the richer groups do not feel any alienation to the poorer groups; if $\theta = 1/2$ then the degree of alienation is symmetric to the groups of higher income and of the lower income, which is also the case of the existing polarization indices and income inequality measures. Different from the existing income polarization indices, therefore, the polarization index $\mathcal{P}(\alpha, \theta)$ reflects not only the distance between the average income levels but also the asymmetric antagonism to the other income-level groups. Figure 1 depicts $\rho_\theta\left(\mu_k - \mu_j\right)$ and the absolute value of the slop determines the degree of the asymmetric alienations of group k to different income-level groups.

The polarization index $\mathcal{P}(\alpha, \theta)$ in (4) depends on the unknown values of π_k's and μ_k's. Therefore, we can obtain an estimator for $\mathcal{P}(\alpha, \theta)$ using proper estimators for π_k's and μ_k's as

$$\widehat{\mathcal{P}}(\alpha, \theta) = \frac{1}{2\overline{y}} \sum_{k=1}^{K} \sum_{j=1}^{K} p_k^{1+\alpha} p_j \rho_\theta\left(\overline{y}_k - \overline{y}_j\right), \tag{5}$$

where $\overline{y} = (1/n) \sum_{i=1}^{n} y_i$, and p_k, p_j, $\overline{y}_k$ and $\overline{y}_j$ are defined as in (1) and (2). Since K is fixed, we can assume that $n_k \to \infty$ for all $k = 1, 2, \cdots, K$ as $n \to \infty$ without loss of

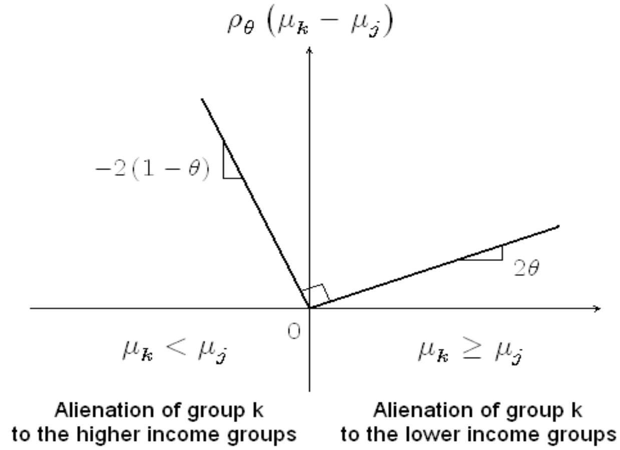

Figure 1: $\rho_\theta\left(\mu_k - \mu_j\right)$ describes asymmetric alienations of group k

generality. Therefore, the consistency of $\widehat{\mathcal{P}}(\alpha, \theta)$ to the true polarization index $\mathcal{P}(\alpha, \theta)$ readily follows by applying the continuous mapping theorem because, for an *i.i.d.* sample, the sample proportion and the sample mean are consistent estimators for the population fraction and the population mean respectively (i.e., $p_k \rightarrow_p \pi_k$ and $\overline{y}_k \rightarrow_p \mu_k$ for each k as $n_k \rightarrow \infty$, where "$\rightarrow_p$" signifies convergence in probability). In the Appendix, further asymptotic properties of $\widehat{\mathcal{P}}(\alpha, \theta)$, including its asymptotic distribution, are provided.

As a special example, we can consider the bipolarization of the income distribution (i.e., $K = 2$). In this case, the polarization index $\mathcal{P}(\alpha, \theta)$ can be simplified as[5]

$$\begin{aligned} \mathcal{B}(\alpha, \theta) &= \frac{1}{2\mu}\left\{\pi_1^{1+\alpha}\pi_2\rho_\theta\left(\mu_1 - \mu_2\right) + \pi_1\pi_2^{1+\alpha}\rho_\theta\left(\mu_2 - \mu_1\right)\right\} \\ &= \frac{\mu_2 - \mu_1}{\mu}\pi_1\pi_2\left[(1-\theta)\pi_1^\alpha + \theta\pi_2^\alpha\right], \end{aligned} \quad (6)$$

where $\pi_2 = 1 - \pi_1$. Notice that, when $\theta = 1/2$, the bipolarization index is reduced to $\mathcal{B}(\alpha, 1/2) = \pi_1\pi_2\left(\pi_1^\alpha + \pi_2^\alpha\right)\left(\mu_2 - \mu_1\right)/2\mu$, which is 1/2 of what Esteban and Ray (1994)

[5]This is because $\rho_\theta\left(\mu_1 - \mu_2\right) = 2\left(\mu_1 - \mu_2\right)(\theta - 1)$ and $\rho_\theta\left(\mu_2 - \mu_1\right) = 2\left(\mu_2 - \mu_1\right)\theta$ for $\mu_1 < \mu_2$. We can write it more simply as

$$\mathcal{B}(\alpha, \theta) = \left(1 - \frac{\mu_1}{\mu}\right)\pi_1\left[(1-\theta)\pi_1^\alpha + \theta\left(1 - \pi_1\right)^\alpha\right]$$

since $\pi_2 = 1 - \pi_1$ and $\mu = \mu_1\pi_1 + \mu_2\pi_2$.

proposed. Similarly as for $\mathcal{P}(\alpha, \theta)$, a consistent estimator for $\mathcal{B}(\alpha, \theta)$ can be obtained as

$$\widehat{\mathcal{B}}(\alpha, \theta) = \frac{\overline{y}_2 - \overline{y}_1}{\overline{y}} p_1 p_2 \left[(1-\theta) p_1^{\alpha} + \theta p_2^{\alpha} \right]$$

for given (α, θ).

3.3 Generalized Polarization Index

The crucial assumption of Theorem 1 is that the intervals $A_k = [a_{k-1}, a_k)$ for $k = 1, 2, \cdots, K$ are pre-determined. Empirically, the choice of the intervals A_k is not easy because even when the number of intervals K is chosen, the location of the cutoff points $\{a_k\}_{k=1}^{K-1}$ are still undetermined. To solve this problem, Esteban, Gradin and Ray (1999) employ Aghevli and Mehran (1981)'s method of optimal grouping when K is given.[6] The idea is that one minimizes the sum of within-group income inequalities (or income dispersion), which are measured by the unnormalized Gini coefficient (i.e., mean difference) for each group, with respect to the optimal cutoff points. Geometrically, this method corresponds to approximating the continuous Lorenz curve by piecewise linear functions and finding the optimal cutoff points minimizing the overall approximation error. More precisely, we select the cutoff points by solving the minimization problem given by[7]

$$\left(a_1^*, \cdots, a_{K-1}^*\right) = \underset{a_1, \cdots, a_{K-1}}{\arg\min}\, \varepsilon\left(a_1, \cdots, a_{K-1}\right) = \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} \int_{a_{k-1}}^{a_k} |x - y| \, dF(x) \, dF(y), \quad (7)$$

where $y_{\min} = a_0^* < a_1^* < \cdots < a_{K-1}^* < a_K^* = y_{\max}$ is satisfied. When we consider the bipolarization index, the number of intervals is already fixed as two $(K = 2)$ and we only need to choose one cutoff point $a_1^* (\equiv y^*)$, which separates the entire income distribution into two. Aghevli and Mehran (1981) show that the optimal cutoff point is easily obtained by the population mean (i.e., $y^* = \mu$) when $K = 2$.

---

[6] More generally, we can use cluster analysis to calssify the income data set into K groupst such as a K-means clustering algorithm (e.g., Hartigan and Wong, 1979).

[7] Equivalently we can minimize $\int_{y_{\min}}^{y_{\max}} \int_{y_{\min}}^{y_{\max}} |x - y| \, dF(x) \, dF(y) - \sum_{k=1}^{K} \sum_{j=1}^{K} \pi_k \pi_j \left| \mu_k - \mu_j \right|$ with respect to $\{a_k\}_{k=1}^{K-1}$.

In addition to selecting the cutoff points, Esteban, Gradin and Ray (1999) use the grouping error $\varepsilon\left(a_1^*, \cdots, a_{K-1}^*\right)$ to generalize the polarization index in that

$$\begin{aligned} EGR_K(\alpha,\beta) &= ER_K^*(\alpha) - \frac{\beta}{2\mu}\varepsilon\left(a_1^*, \cdots, a_{K-1}^*\right) \\ &= \frac{1}{\mu}\sum_{k=1}^{K}\sum_{j=1}^{K}\left(\pi_k^*\right)^{1+\alpha}\pi_j^*\left|\mu_k^* - \mu_j^*\right| - \frac{\beta}{2\mu}\varepsilon\left(a_1^*, \cdots, a_{K-1}^*\right) \end{aligned}$$

with some weight $\beta > 0$, where π_k^* and μ_k^* are obtained over the intervals $A_k^* = [a_{k-1}^*, a_k^*)$ defined by the optimal cutoff points $\{a_k^*\}_{k=1}^{K-1}$ for all $k = 1, 2, \cdots, K$. The motivation is that the original polarization index $ER_K(\alpha)$ over-estimates the within-group identification by construction and it can be adjusted by subtracting the approximation error $\varepsilon\left(a_1^*, \cdots, a_{K-1}^*\right)$ with some weight β. For example, when we calculate the bipolarization index $\mathcal{B}(\alpha,\theta)$, we consider $Bernoulli(\pi_1)$ distribution over the support points μ_1 and μ_2 instead of the original distribution $F(y)$ of $\{y_i\}_{i=1}^n$. By doing so, we force all the income levels to be concentrated on each group mean (i.e., either μ_1 or μ_2), which is the extreme case of the group identification. Since $\varepsilon\left(a_1^*, \cdots, a_{K-1}^*\right)$ measures the difference (or the approximation error) between the continuous distribution $F(y)$ and the discrete representation, therefore, subtracting the error term can be a remedy for such over-estimation.[8]

We can still simply extend the polarization index $\mathcal{P}(\alpha,\theta)$ similarly as Esteban, Gradin and Ray (1999).[9] One major drawback of such generalization is that, however, we need

[8] The logic is that as each group presents more identity (i.e., less income dispersion within the group), then the error $\varepsilon\left(a_1^*, \cdots, a_{K-1}^*\right)$ becomes smaller and thus the polarization index $EGR_K(\alpha,\beta)$ gets larger. On the other hand, if the group identity gets weaker (i.e., more income dispersion within the group), then $EGR_K(\alpha,\beta)$ becomes smaller.

[9] For example, using the same argument of Esteban, Gradin and Ray (1999), we can also generalize the bipolarization index $\mathcal{B}(\alpha,\theta)$ as

$$\begin{aligned} \mathcal{B}_G(\alpha,\theta,\beta) &= \mathcal{B}^*(\alpha,\theta) - \beta\varepsilon(y^*)/2\mu \qquad (8) \\ &= \left\{(1-\theta)\pi_*^\alpha + \theta(1-\pi_*)^\alpha\right\}(\pi_* - L(\pi_*)) - \beta\left\{\frac{1}{2\mu}\int_{y_{\min}}^{y_{\max}}\int_{y_{\min}}^{y_{\max}}|x-y|\,dF(x)\,dF(y) - (\pi_* - L(\pi_*))\right\} \\ &= \left\{(1-\theta)\pi_*^\alpha + \theta(1-\pi_*)^\alpha + \beta\right\}(\pi_* - L(\pi_*)) - \frac{\beta}{2\mu}\int_{y_{\min}}^{y_{\max}}\int_{y_{\min}}^{y_{\max}}|x-y|\,dF(x)\,dF(y), \end{aligned}$$

where y^* is the optimal cutoff point and $\pi_* = \int_{y_{\min}}^{y^*=\mu} dF(y)$. Such expression uses the argument that $\varepsilon(y^*)$ corresponds to the approximation error between the continuous Lorenz curve and its piecewise linear to decide a new weight value β. There is no guidance to choose a proper β and therefore, we can have any value of $EGR_K(\alpha, \beta)$ depending on the choice of β, even negative. Such an arbitrariness also affects the variance of $EGR_K(\alpha, \beta)$ and its t-value. Alternatively, we consider a different form of the generalized (or bias corrected) polarization index given by

$$\mathcal{P}_G(\alpha, \theta) = \frac{1}{2\mu} \sum_{k=1}^{K} \sum_{j=1}^{K} \pi_k \pi_j \left(\frac{\pi_k}{G_k/G} \right)^{\alpha} \rho_\theta (\mu_k - \mu_j)$$

and

$$\begin{aligned} \mathcal{B}_G(\alpha, \theta) &= \frac{\mu_2 - \mu_1}{\mu} \pi_1 \pi_2 \left[(1-\theta) \left(\frac{\pi_1}{G_1/G} \right)^{\alpha} + \theta \left(\frac{\pi_2}{G_2/G} \right)^{\alpha} \right] \\ &= \frac{\mu_2 - \mu_1}{\mu} \pi_1 \pi_2 G^{\alpha} \left[(1-\theta) \left(\frac{\pi_1}{G_1} \right)^{\alpha} + \theta \left(\frac{\pi_2}{G_2} \right)^{\alpha} \right] \\ &= \left(1 - \frac{\mu_1}{\mu} \right) \pi_1 G^{\alpha} \left[(1-\theta) \left(\frac{\pi_1}{G_1} \right)^{\alpha} + \theta \left(\frac{1-\pi_1}{G_2} \right)^{\alpha} \right], \end{aligned}$$

where G is the Gini index of the entire sample and G_k is the Gini index of the income distribution over the interval A_k. In this specification, the within-group identity is measured by the term $(\pi_k / (G_k/G))^{\alpha}$, which gets larger as the income levels within group k become more identical. This measures *relative* dispersion of the within-group income distribution, which can be more meaningful than the absolute dispersion measure (e.g., $(\pi_k/G_k)^{\alpha}$) since changes in G_k also alter the overall income inequality G. Using the similar argument as for

---

approximation function. Note that the Lorenz curve satisfies

$$L(\pi_*) = \frac{1}{\mu} \int_0^{\pi_*} F^{-1}(t)\, dt = \frac{1}{\mu} \int_{y_{\min}}^{y^*=\mu} y dF(y) = \frac{1}{\mu} \pi_* \mu_1^*$$

by change of variables, where $\mu_1^* = \mathbb{E}[y | y \in A_1^*]$. In addition, since we have $\mu_1^* = \mu L(\pi_*)/\pi_*$ and $\mu_2^* = \mu (1 - L(\pi_*)) / (1 - \pi_*)$, we can rewrite (8) as

$$\mathcal{B}_G(\alpha, \theta, \beta) = \pi_* \left(1 - \frac{\mu_1^*}{\mu} \right) \{ (1-\theta) \pi_*^{\alpha} + \theta (1 - \pi_*)^{\alpha} + \beta \} - \frac{\beta}{2\mu} \int_{y_{\min}}^{y_{\max}} \int_{y_{\min}}^{y_{\max}} |x - y|\, dF(x)\, dF(y),$$

which can be consistently estimated by

$$\widehat{\mathcal{B}}_G(\alpha, \theta, \beta) = p_* \left(1 - \frac{\overline{y}_1^*}{\overline{y}} \right) \{ (1-\theta) p_*^{\alpha} + \theta (1 - p_*)^{\alpha} + \beta \} - \frac{\beta}{2\overline{y} n^2} \sum_{i=1}^{n} \sum_{h=1}^{n} |y_i - y_h|, \qquad (9)$$

where $p_* = n_1^*/n$ and $\overline{y}_1^* = (1/n_1^*) \sum_{i=1}^{n} y_i \mathbb{I}\{y_i \in A_1^*\}$.

$EGR_K(\alpha,\beta)$, $(G_k/G)^{-\alpha}$ term can be interpreted as a correction term for the over-estimated within-group identification. Note that $\mathcal{P}_G(\alpha,\theta)$ is more convenient than $EGR_K(\alpha,\beta)$ since it does not introduce further arbitrary parameter β.[10] Similarly as in the previous subsection, $\mathcal{P}_G(\alpha,\theta)$ and $\mathcal{B}_G(\alpha,\theta)$ can be consistently estimated as follows:

$$\widehat{\mathcal{P}}_G(\alpha,\theta) = \frac{1}{2\overline{y}}\sum_{k=1}^{K}\sum_{j=1}^{K} p_k p_j \left(\frac{p_k}{\widehat{G}_k/\widehat{G}}\right)^{\alpha} \rho_\theta\left(\overline{y}_i - \overline{y}_j\right), \tag{10}$$

$$\widehat{\mathcal{B}}_G(\alpha,\theta) = \frac{\overline{y}_2 - \overline{y}_1}{\overline{y}} p_1 p_2 \left[(1-\theta)\left(\frac{p_1}{\widehat{G}_1/\widehat{G}}\right)^{\alpha} + \theta\left(\frac{p_2}{\widehat{G}_2/\widehat{G}}\right)^{\alpha}\right], \tag{11}$$

where $\widehat{G}_k = (1/2\overline{y}_k n_k(n_k-1))\sum\sum_{i\neq j}|y_i - y_j|\,\mathbb{I}\{y_i \in A_k\}\,\mathbb{I}\{y_j \in A_k\}$ is the Gini coefficient for group k and $\widehat{G} = (1/2\overline{y}n(n-1))\sum\sum_{i\neq j}|y_i - y_j|$ is the conventional Gini coefficient.[11]

4 Empirical Results *(preliminary)*

4.1 Model

In our crime model, the logarithm of the crime rate, property or violent, is explained by the Gini coefficient and the bipolarization index along with other control variables, which include two crime deterrent variables (the number of police officers per capita and the arrest rate), the proportion of population who are men and between 15 and 29, and the unemployment rate. The primary goal is to test if, holding all the other variables constant including the degree of income inequality, crime rates increase as the income distribution gets more bipolarized.

In the estimation, the simultaneity between the crime rate and crime deterrent variables,

[10] Similarly as $\mathcal{P}(\alpha,\theta)$, it should be also noted that the generalized polarization index $\mathcal{P}_G(\alpha,\theta)$ still satisfy the axioms of Duclos, Esteban and Ray (2004). In this case, we simply define the identity function as $\mathcal{I}(x) = (\pi_k/(G_k/G))\,\mathbb{I}\{x \in A_k\}$.

[11] Asymptotic distribution of $\widehat{\mathcal{P}}_G(\alpha,\theta)$ or $\widehat{\mathcal{B}}_G(\alpha,\theta)$ are hard to obtain since the asymptotic variance formula of the Gini coefficient estimator is very complicated and hard to estimate; and it is even harder to obtain the covariance terms between the Gini coefficient estimator and $(p_k,\overline{y}_k)_{k=1,\cdots,K}$. Even when we could find the analytical asymptotic variance formula of them, the feasibility of estimating this variance is still questionable. So, in Appendix, we instead suggest a Jackknife variance estimator of $\widehat{\mathcal{B}}_G(\alpha,\theta)$.

the police size and the arrest rate, is addressed by using instrumental variables that are correlated with deterrent variables and uncorrelated with the crime rate. In the property crime model, per-capita tax revenue, the arrest rate of violent crime, and weighted average of per-capita police sizes of neighboring states are used as instrumental variables for the police size and the arrest rate of property crime. In the violent crime model, per-capita tax revenue, the arrest rate of property crime, and weighted average of per-capita police sizes of neighboring states are used as instrumental variables for the police size and the arrest rate of violent crime. Unobservable state-specific fixed effects are included and fixed-effect Generalized Method of Moments (GMM) estimation is applied with HAC robust standard error estimates.

4.2 Data

We use the U.S. household-level income data from PSID to construct state-level bipolarization index. As discussed by Kelly (2000) and Demombynes and Ozler (2005) among others, an appropriate unit of observation for which crime is examined might well be smaller than the state level. Two difficulties arise, however, when we adopt a very small observational unit. First, variables such as income inequality are not readily available at aggregation levels lower than the state level. If one tries to estimate such indices as the Gini coefficient and bipolarization index based on individual household data, not enough observations are available at the county or police district level.[12] Second, when the observation unit is very small, the local crime rate does not necessarily reflect the region's economic conditions, as criminals travel to neighborhoods in searching for higher returns (Demombynes and Olzer, 2005) or those who are frustrated in one region move out to another region where they have better prospects and therefore decide to supply labor to the legal labor market. These latter two possibilities tend to make the analysis extremely complicated.

The problem of openness of the crime market can be avoided by adopting a large obser-

[12] That may explain why, to the best of my knowledge, none of the papers that adopted a observational unit smaller than or equal to the county level such as Cornwell and Trumbell (1994), Glaeser, Sacerdote, and Scheinkman (1996), Wilson and Daly (1997), and Kelly (2000) addressed the issue of the relationship between inequality and crime.

vational unit such as country. Cross-country comparison of crime rates, however, is often obscured by differences across countries in the definition of certain crimes and the degree of under-reporting (Fajnzylber, Lederman, and Loayza, 2002). It also suffers from misalignment of data for various countries for the same time period. More importantly, a person's crime behaviors are influenced by the local market conditions of the area within which the person's economic activities can actually be pursued.

All these problems could be minimized by using the state level data. To summarize, crime markets are relatively closed at the state level so that only own-state characteristics are relevant for the state's crime rate, definitions of various crime categories are consistent across states, a large number of individual or household observations are available for most states to compute various measures of inequality and bipolarization, and finally balanced panel data are available for a relatively long period of time.

The crime data are directly drawn from the FBI Uniform Crime Reports for the period of 1991 through 2005. We analyze both violent and property crime rates, which are defined as the number of reported offenses per 100,000 population.[13] We use the Current Population Survey data for the same sample period to compute the Gini coefficient and the bipolarization indices by state and by year. The final sample includes 600 state-year observations that have valid information on all the variables used in our regression. The data is balanced in the sense that each state has the entire 15 years of observations.[14]

4.3 Findings

Table I shows the preliminary results for property and violent crime. When an equal weight is placed on both the high and low income groups ($\theta = 0.5$; the first columns), the coefficient of our bipolarization measure is estimated insignificantly. When greater weights are applied to the low income group ($\theta > 0.5$; the second and the third columns), the coefficient of the bipolarization index becomes positive and statistically significant even at the one percent

[13] For the 1991 through 1993 period, the two crime deterrent variables are directly obtained from the Federal Bureau of Investigation.

[14] Exclusion of 11 states is mainly due to missing observations on the arrest variable.

level. This finding supports our asymmetry hypothesis that individuals feel more frustrated (and thus more alienation) when they belong to the low income group than high income group. In our income mobility interpretation, given that enhanced mobility between groups mitigates the antagonism that the low-income earners have and increases their feeling of frustration to the high-income group, the former effect is greater than the latter effect in an absolute value in explaining the crime rate. The estimated coefficient of the Gini index is negative and significant in our preferred models.

TABLE I: Fixed-Effect GMM Estimates[a]

	Property Crime			Violent Crime		
	$\theta = .5$	$\theta = .25$	$\theta = 0$	$\theta = .5$	$\theta = .25$	$\theta = 0$
Gini	-0.166	-1.480	-1.622	-0.397	-1.354	-1.435
	(0.526)	(0.451)	(0.421)	(0.702)	(0.595)	(0.552)
Bipolarization	-0.611	0.997	0.810	-0.264	0.806	0.629
	(0.731)	(0.361)	(0.228)	(0.995)	(0.482)	(0.304)
Police	-6.652	-6.761	-6.791	-8.042	-8.053	-8.073
	(0.855)	(0.851)	(0.852)	(1.134)	(1.134)	(1.136)
Arrest	1.075	1.063	1.063	0.285	0.266	0.266
	(0.322)	(0.316)	(0.314)	(0.207)	(0.205)	(0.205)
Young Men	0.869	0.787	0.814	2.718	2.636	2.612
	(2.100)	(2.058)	(2.039)	(2.960)	(2.941)	(2.935)
Unemployment	-0.098	0.266	0.441	-0.767	-0.449	-0.325
	(0.723)	(0.716)	(0.714)	(0.946)	(0.941)	(0.943)
N. of State		40			40	
obs.		600			600	

a: Robust standard error estimates are in parentheses. α=1.6.

5 Concluding Remarks

We derive a testable hypothesis regarding the relationship between income bipolarization and crime. As an income distribution is more bipolarized, mobility between different income groups is reduced, which lowers the crime incentive of the richer, but raise that of the poorer.

We assume that the latter effect dominates the former, producing a positive association of the degree of bipolarization and the crime rate.

For the purpose of testing our hypothesis, the bipolarization measure developed by Esteban, Gradin, and Ray (1999) is generalized by allowing asymmetric degrees of alienation between the low- and the high-income classes. Our fixed effects GMM estimates based on state level data for United States from 1991 through 2005 suggests that, while the Gini index (the income inequality measure) is not related with any type of crime, bipolarization significantly increases both the property and the violent crime rates. More importantly, consistent with our theoretical prediction regarding asymmetric alienation feeling between the rich and the poor, the explanatory power of our generalized polarization index increases as we put a greater weight on the lower income class.

Appendix: Technical Results

A.1 Asymptotic Distribution

The asymptotic normality of $\widehat{\mathcal{P}}(\alpha,\theta)$ can be derived as follows. "$\rightarrow_d$" signifies convergence in distribution.

Theorem A1 (Polarization Index) *We assume that $\pi_k > 0$ and $n_k \rightarrow \infty$ as $n \rightarrow \infty$ for all $k = 1, 2, \cdots, K$. Then as $n \rightarrow \infty$,*

$$\sqrt{n}\left(\widehat{\mathcal{P}}(\alpha,\theta) - \mathcal{P}(\alpha,\theta)\right) \rightarrow_d \mathcal{N}\left(0, \sigma^2_{\mathcal{P}}(\alpha,\theta)\right) \tag{12}$$

for given α and θ, where the asymptotic variance $\sigma^2_{\mathcal{P}}(\alpha,\theta)$ is defined as

$$\sigma^2_{\mathcal{P}}(\alpha,\theta) = \sum_{k=1}^{K}\left(\gamma^2_{K,k}(\alpha,\theta)\,\pi_k + \frac{\delta^2_{K,k}(\alpha,\theta)\,\sigma^2_k}{\pi_k}\right) - (\alpha+1)^2\,\mathcal{P}^2(\alpha,\theta)$$

with

$$\gamma_{K,k}(\alpha,\theta)=\frac{1}{2\mu}\sum_{j=1,j\neq k}^{K}\pi_j\left[(1+\alpha)\,\pi_k^{\alpha}\rho_\theta\left(\mu_k-\mu_j\right)+\pi_j^{\alpha}\rho_\theta\left(\mu_j-\mu_k\right)\right]-\frac{\mu_k}{\mu}\mathcal{P}(\alpha,\theta) \quad (13)$$

$$\delta_{K,k}(\alpha,\theta)=\frac{\pi_k}{\mu}\left\{\sum_{j=1}^{k-1}\pi_j\left(\pi_k^{\alpha}+\pi_j^{\alpha}\right)-\sum_{j=1,j\neq k}^{K}\pi_j\left[(1-\theta)\,\pi_k^{\alpha}+\theta\pi_j^{\alpha}\right]-\mathcal{P}(\alpha,\theta)\right\}. \quad (14)$$

Proof of Theorem A1 From Gastwirth, Nayak and Kreiger (1986) or Sethuraman (1963), we have

$$\sqrt{n}\begin{pmatrix} p_1-\pi_1 \\ \vdots \\ p_K-\pi_K \\ \overline{y}_1-\mu_1 \\ \vdots \\ \overline{y}_K-\mu_K \end{pmatrix}\rightarrow_d \mathcal{N}\left(0,\begin{pmatrix}\Sigma_\pi & 0 \\ 0 & \Sigma_\mu\end{pmatrix}\right) \quad (15)$$

as $n\rightarrow\infty$, where the $K\times K$ matrices Σ_π and Σ_μ are defined as

$$\Sigma_\pi=\begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_K \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_K \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_1\pi_K & -\pi_2\pi_K & \cdots & \pi_K(1-\pi_K) \end{pmatrix} \quad \textit{and} \quad \Sigma_\mu=diag\left(\frac{\sigma_1^2}{\pi_1},\cdots,\frac{\sigma_K^2}{\pi_K}\right).$$

Now the basic strategy of the proof is similar to Gastwirth, Nayak and Kreiger (1986). For notational simplicity, we omit the argument (α,θ) of $\mathcal{P}$, $\gamma_{K,k}$, $\delta_{K,k}$ and $\sigma_{\mathcal{P}}^2$. We can show that, for given true values of $(\pi_k,\mu_k)_{k=1}^K$, $\partial\mathcal{P}/\partial\pi_k=\gamma_{K,k}$ and $\partial\mathcal{P}/\partial\mu_k=\delta_{K,k}$, where $\gamma_{K,k}$ and $\delta_{K,k}$ are defined as in (13) and (14) as functions of $(\pi_k,\mu_k)_{k=1}^K$.

We let $\bigtriangledown\mathcal{P}=(\partial\mathcal{P}/\partial\pi_1,\cdots,\partial\mathcal{P}/\partial\pi_K,\partial\mathcal{P}/\partial\mu_1,\cdots,\partial\mathcal{P}/\partial\mu_K)'$, which is evaluated at the true parameter values $\left(\pi_k,\mu_k,\sigma_k^2\right)_{k=1}^K$. Using (15) and the delta-method, we then have the asymptotic normality result (12), where the asymptotic variance is obtained by

$$\begin{aligned}
\sigma^2_{\mathcal{P}} &= \bigtriangledown \mathcal{P}' \begin{pmatrix} \Sigma_\pi & 0 \\ 0 & \Sigma_\mu \end{pmatrix} \bigtriangledown \mathcal{P} \\
&= \sum_{k=1}^{K} \gamma^2_{K,k}\pi_k - \left(\sum_{k=1}^{K} \gamma_{K,k}\pi_k\right)^2 + \sum_{k=1}^{K} \frac{\delta^2_{K,k}\sigma^2_k}{\pi_k} \\
&= \sum_{k=1}^{K} \left(\gamma^2_{K,k}(\alpha,\theta)\pi_k + \frac{\delta^2_{K,k}(\alpha,\theta)\sigma^2_k}{\pi_k}\right) - (\alpha+1)^2 \mathcal{P}^2.
\end{aligned}$$

The last equation is followed from

$$\begin{aligned}
\sum_{k=1}^{K} \gamma_{K,k}\pi_k &= \frac{1}{2\mu} \sum_{k=1}^{K}\sum_{j=1}^{K} \pi_k\pi_j \left[(1+\alpha)\pi_k^\alpha \rho_\theta\left(\mu_k - \mu_j\right) + \pi_j^\alpha \rho_\theta\left(\mu_j - \mu_k\right)\right] - \frac{\sum_{k=1}^{K} \mu_k\pi_k}{\mu}\mathcal{P} \\
&= (1+\alpha)\mathcal{P} + \mathcal{P} - \mathcal{P} = (1+\alpha)\mathcal{P}
\end{aligned}$$

since $\mu = \sum_{i=1}^{K} \mu_i\pi_i$. *Q.E.D.*

The asymptotic variance $\sigma^2_{\mathcal{P}}(\alpha,\theta)$ can be consistently estimated by

$$\widehat{\sigma}^2_{\mathcal{P}}(\alpha,\theta) = \sum_{k=1}^{K} \left(\widehat{\gamma}^2_{K,k}(\alpha,\theta) p_k + \frac{\widehat{\delta}^2_{K,k}(\alpha,\theta) s^2_k}{p_k}\right) - \left(\sum_{k=1}^{K} \widehat{\gamma}_{K,k}(\alpha,\theta) p_k\right)^2,$$

where

$$\begin{aligned}
\widehat{\gamma}_{K,k}(\alpha,\theta) &= \frac{1}{2\overline{y}} \sum_{j=1}^{K} p_j \left[(1+\alpha) p_k^\alpha \rho_\theta\left(\overline{y}_k - \overline{y}_j\right) + p_j^\alpha \rho_\theta\left(\overline{y}_j - \overline{y}_k\right)\right] - \frac{\overline{y}_k}{\overline{y}}\widehat{\mathcal{P}}(\alpha,\theta) \quad \text{and} \\
\widehat{\delta}_{K,k}(\alpha,\theta) &= \frac{p_k}{\overline{y}} \left\{ \sum_{j=1}^{k-1} p_j\left(p_k^\alpha + p_j^\alpha\right) - \sum_{j=1, j\neq k}^{K} p_j\left[(1-\theta)p_k^\alpha + \theta p_j^\alpha\right] - \widehat{\mathcal{P}}(\alpha,\theta) \right\}.
\end{aligned}$$

Note that $p_k \to_p \pi_k$, $\overline{y}_k \to_p \mu_k$ and $s^2_k \to_p \sigma^2_k$ for each k ensure that $\widehat{\gamma}_{K,k}(\alpha,\theta) \to_p \gamma_{K,k}(\alpha,\theta)$ and $\widehat{\delta}_{K,k}(\alpha,\theta) \to_p \delta_{K,k}(\alpha,\theta)$ as $n \to \infty$ for given θ and α. Note that when $\theta = 1/2$, Theorem 1 also provides the the asymptotic normality of the polarization index estimator for $ER_K(\alpha) = (1/\mu)\sum_{k=1}^{K}\sum_{j=1}^{K} \pi_k^{1+\alpha}\pi_j \left|\mu_k - \mu_j\right|$, which is suggested by Esteban and Ray

(1994) since $ER_K(\alpha) = 2\mathcal{P}(\alpha, 1/2)$. Using Theorem 1, the asymptotic distribution of $\widehat{\mathcal{B}}(\alpha, \theta)$ is summarized as follows.

Corollary A2 (Bipolarization Index) *When $K = 2$, under the same conditions as Theorem 1 and as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\mathcal{B}}(\alpha,\theta) - \mathcal{B}(\alpha,\theta)\right) \to_d \mathcal{N}\left(0, \sigma^2_{\mathcal{B}}(\alpha,\theta)\right)$$

for given α and θ, where the asymptotic variance $\sigma^2_{\mathcal{B}}(\alpha,\theta)$ is given by

$$\sigma^2_{\mathcal{B}}(\alpha,\theta) = \sum_{k=1}^{2}\left(\gamma_k^2(\alpha,\theta)\pi_k + \frac{\delta_k^2(\alpha,\theta)\sigma_k^2}{\pi_k}\right) - (\alpha+1)^2\mathcal{B}^2(\alpha,\theta)$$

with

$$\gamma_1(\alpha,\theta) = \left(\frac{\mu_2-\mu_1}{\mu}\right)\pi_2\left[(1+\alpha)(1-\theta)\pi_1^\alpha + \theta\pi_2^\alpha\right] - \frac{\mu_1}{\mu}\mathcal{B}(\alpha,\theta),$$
$$\gamma_2(\alpha,\theta) = \left(\frac{\mu_2-\mu_1}{\mu}\right)\pi_1\left[(1-\theta)\pi_1^\alpha + (1+\alpha)\theta\pi_2^\alpha\right]\theta - \frac{\mu_2}{\mu}\mathcal{B}(\alpha,\theta),$$

and

$$\delta_1(\alpha,\theta) = -\frac{\pi_1\pi_2}{\mu}\left[(1-\theta)\pi_1^\alpha + \theta\pi_2^\alpha\right] - \frac{\pi_1}{\mu}\mathcal{B}(\alpha,\theta),$$
$$\delta_2(\alpha,\theta) = \frac{\pi_1\pi_2}{\mu}\left[(1-\theta)\pi_1^\alpha + \theta\pi_2^\alpha\right] - \frac{\pi_2}{\mu}\mathcal{B}(\alpha,\theta).$$

The asymptotic variance $\sigma^2_{\mathcal{B}}(\alpha,\theta)$ can be also consistently estimated using $\overline{y}_k$, p_k and s_k^2 for $k = 1, 2$.

A.2 Cutoff Point Selection When $K = 2$

When $K = 2$, we can consider a different method in choosing the cutoff point. Since the Lorenz curve is a function of the distribution function, one can instead choose the cutoff point by minimizing the approximation error between true distribution function $F(y)$ and its piecewise linear approximation directly.

Statistically, this approach is equivalent to minimizing the error between the true density function $f(y) = dF(y)/dy$ and its piecewise constant approximation, i.e., the histogram estimator. Especially for the bipolarization index case, if we consider the L_2-distance (e.g., we consider the integrated mean squared error as the distance between the histogram estimator and the true density function), this problem can be considered as the bandwidth parameter selection problem. Notice that Aghevli and Mehran (1981) minimize the L_1-distance (i.e., the mean difference), on the other hand.

More precisely, we consider a histogram estimator given by

$$\widehat{f}_*(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_k}\mathbb{I}\{y_i \in A_k^*\} \quad \text{for } y \in A_k^* \text{ and } k = 1, 2,$$

where $A_1^* = [y_{\min}, y^*)$ and $A_2^* = [y^*, y_{\max}]$ with $y_{\min} < y^* < y_{\max}$; h_k is the length of each interval (i.e., $h_1 = h_1(y^*) = y^* - y_{\min}$ and $h_2 = h_2(y^*) = y_{\max} - y^*$). Different from the conventional histogram estimator, in this case, we let the number of intervals fixed as two and the curve fit is optimized by choosing the value y^*, which splits the support $[y_{\min}, y_{\max}]$ into two. In other words, we control the curve fit by choosing the optimal bandwidth parameter, which could be varying.

Since we fix the number of intervals ($K = 2$) in this case, h_k cannot go to zero as the sample size n increases. It certainly violates the standard assumption in kernel density estimator (i.e., $h \to 0$ and $nh \to \infty$ as $n \to \infty$), and we thus cannot show the consistency of $\widehat{f}_*(y)$ for the true density function $f(y)$. We can still, however, choose y^*, which minimizes the integrated mean squared error (MISE) of $\widehat{f}_*$ given by

$$\begin{aligned} MISE\left(\widehat{f}_*\right) &= \frac{n+1}{n}\sum_{k=1}^{2}\left(f'(m_k)\right)^2 h_k\left\{\frac{1}{3}\left(a_{k-1}^2 + a_k a_{k-1} + a_k^2\right) - m_k^2\right\} \\ &\quad + \frac{2}{n}\sum_{k=1}^{2} f'(m_k)\{m_k\pi_k - \mu_k\} + \frac{1}{n}\sum_{k=1}^{2}\frac{\pi_k}{h_k} + \frac{1}{n}\|f\|_2^2, \end{aligned}$$

where $\|f\|_2^2 = \int f^2(y)\, dy$ and m_k is the mid-point of each interval. The derivation is provided in Lemmas A3 and A4 below. As in the standard kernel density estimation, however, we cannot use this method immediately to find y^* since the true density function f is unknown. Several methods has been proposed in the nonparametric literature such as Silverman's rule of thumb or cross validation, and we can also apply these methods in this context.

Lemma A3 (Bias and Variance of $\widehat{f}_*(y)$) *Assume that $f(y)$ is bounded and twice continuously differentiable with bounded second-order derivatives around y, where y is an interior point of the support. Then, for $y \in A_k$ $(k = 1, 2, \cdots, K)$,*

$$Bias\left(\widehat{f}_*(y)\right) = f'(m_k)(m_k - y)$$

and

$$var\left(\widehat{f}_*(y)\right) = \frac{1}{nh_k}\left\{f(y) + f'(m_k)(m_k - y)\right\} - \frac{1}{n}\left\{f(y) + f'(m_k)(m_k - y)\right\}^2,$$

where m_k is the mid-point of each bin A_k, i.e., $m_k = (a_k + a_{k-1})/2$.

Notice that the variance approximation of Lemma A2 is somewhat different from the conventional formula of the histogram estimator. It is regularly presumed that $h \to 0$ as $n \to \infty$ in the conventional context and thus the first term of $O\left(n^{-1}h^{-1}\right)$ dominates the second term of $O\left(n^{-1}\right)$. For the case of fixed number of intervals, however, the varying bandwidth h_k does not vanish as $n \to \infty$ and we cannot rely on the standard argument in nonparametric density estimation.

Proof of Lemma A3 For the bias, we first note that the first-order Taylor approximation of $f(u) - f(y)$ around the center of each interval A_k yields $f(u) - f(y) \simeq f'(m_k)(u - y)$, where $f'(y) = df(y)/dy$ and m_k is the mid-point of each bin A_k. Therefore, for $y \in A_k$

$$\begin{aligned}
\mathbb{E}\widehat{f}_*(y) - f(y) &= \frac{1}{h_k}\mathbb{P}\{y_i \in A_k\} - f(y) \\
&= \frac{1}{h_k}\int_{A_k} f(u) - f(y)\, du \\
&\simeq \frac{1}{h_k}\int_{A_k} f'(m_k)(u-y)\, du \\
&= f'(m_k)(m_k - y)
\end{aligned}$$

since $\{y_i\}_{i=1}^n$ are *i.i.d.* and the length of the interval A_k is given as $\int_{A_k} du = h_k$ by construction. For the variance, using the properties of the Bernoulli random variable,

$$\begin{aligned}
var\left(\widehat{f}_*(y)\right) &= \frac{1}{nh_k^2} var\left(\mathbb{I}\{y_i \in A_k\}\right) \\
&= \frac{1}{nh_k^2}\mathbb{P}\{y_i \in A_k\}\left(1 - \mathbb{P}\{y_i \in A_k\}\right) \\
&= \frac{1}{nh_k^2}\int_{A_k} f(u)\, du - \frac{1}{nh_k^2}\left(\int_{A_k} f(u)\, du\right)^2 \\
&= \frac{1}{nh_k}\left\{f(y) + f'(m_k)(m_k - y)\right\} - \frac{1}{n}\left\{f(y) + f'(m_k)(m_k - y)\right\}^2. \quad Q.E.D.
\end{aligned}$$

The following lemma derives the general form of MISE of $\widehat{f}_*$.

Lemma A4 (MISE) *Under the same conditions in Lemma A3,*

$$\begin{aligned}
MISE\left(\widehat{f}_*\right) &= \frac{n+1}{n}\sum_{k=1}^{K}\left(f'(m_k)\right)^2 h_k \left\{\frac{1}{3}\left(a_{k-1}^2 + a_k a_{k-1} + a_k^2\right) - m_k^2\right\} \\
&\quad + \frac{2}{n}\sum_{k=1}^{K} f'(m_k)\left\{m_k \pi_k - \mu_k\right\} + \frac{1}{n}\sum_{k=1}^{K}\frac{\pi_k}{h_k} + \frac{1}{n}\int f^2(y)\, dy.
\end{aligned}$$

Proof of Lemma A4 Under the regularity condition, the MISE of $\widehat{f}_*$ can be obtained from

$$\begin{aligned}
& \int \left\{ bias^2 \left(\widehat{f}_* (y) \right) + var \left(\widehat{f}_* (y) \right) \right\} dy \\
= & \sum_{k=1}^{K} \int_{A_k} \left[f' (m_k) \right]^2 (m_k - y)^2 dy \\
& + \frac{1}{nh_k} \sum_{k=1}^{K} \int_{A_k} \left\{ f (y) + f' (m_k) (m_k - y) \right\} dy - \frac{1}{n} \sum_{k=1}^{K} \int_{A_k} \left\{ f (y) + f' (mk) (m_k - y) \right\}^2 dy
\end{aligned}$$

using Lemma A3. Note that for each k and interval $A_k = [a_{k-1}, a_k)$,

$$\begin{aligned}
\int_{A_k} \left[f' (m_k) \right]^2 (m_k - y)^2 dy &= \left(f' (m_k) \right)^2 \left\{ m_k^2 h_k - 2m_k \int_{A_k} y dy + \int_{A_k} y^2 dy \right\} \\
&= \left(f' (m_k) \right)^2 h_k \left\{ \frac{1}{3} \left(a_{k-1}^2 + a_k a_{k-1} + a_k^2 \right) - m_k^2 \right\},
\end{aligned}$$

$$\frac{1}{nh_k} \int_{A_k} \left\{ f (y) + f' (m_k) (m_k - y) \right\} dy = \frac{1}{nh_k} \left\{ \pi_k + f' (m_k) \left\{ m_k h_k - \int_{A_k} y dy \right\} \right\} = \frac{\pi_k}{nh_k}$$

and

$$\begin{aligned}
\frac{1}{n} \int_{A_k} \left\{ f (y) + f' (m_i) (m_i - y) \right\}^2 dy &= \frac{1}{n} \int_{A_k} f^2 (y) dy + \frac{2}{n} f' (m_k) \left\{ m_k \pi_k - \mu_k \right\} \\
& \quad + \frac{1}{n} \left(f' (m_k) \right)^2 \left\{ m_k^2 h_k - 2m_k \int_{A_k} y dy + \int_{A_k} y^2 dy \right\} \\
&= \frac{1}{n} \int_{A_k} f^2 (y) dy + \frac{2}{n} f' (m_k) \left\{ m_k \pi_k - \mu_k \right\} \\
& \quad + \frac{1}{n} \left(f' (m_k) \right)^2 h_k \left\{ \frac{1}{3} \left(a_{k-1}^2 + a_k a_{k-1} + a_k^2 \right) - m_k^2 \right\}
\end{aligned}$$

since $\int_{A_k} dy = a_k - a_{k-1} = h_k$, $\int_{A_k} f (y) dy = \pi_k$, $\int_{A_k} y f (y) dy = \mu_k$, $\int_{A_k} y dy = m_k h_k$ and $\int_{A_k} y^2 dy = (1/3) \left(a_{k-1}^2 + a_k a_{k-1} + a_k^2 \right) h_k$ by construction. By combining these results, we can find the $MISE \left(\widehat{f}_* \right)$ as desired. *Q.E.D.*

A.3 Jackknife Variance Estimation

In some cases, we need to obtain the standard error of the polarization index. For example, when we want to compare polarization indices between two different groups or test the changes in polarization over time, the standard error is a key ingredient for constructing any test statistics. In Theorem A1 and Corollary A2, we successfully derive the asymptotic variance of the polarization indices $\mathcal{P}(\alpha,\theta)$ and $\mathcal{B}(\alpha,\theta)$. It is, however, very difficult to find the asymptotic variance formula of the generalized polarization index estimator $\widehat{\mathcal{P}}_G(\alpha,\theta)$ in (10) or even its simplest case $\widehat{\mathcal{B}}_G(\alpha,\theta)$ in (11). The main reason is that the asymptotic variance formula of the Gini coefficient estimator is very complicated and hard to estimate; and it is even harder to obtain the covariance terms between the Gini coefficient estimator and $(p_k,\overline{y}_k)_{k=1,\cdots,K}$. Even when we could find the analytical asymptotic variance formula of $\widehat{\mathcal{P}}_G(\alpha,\theta)$ or $\widehat{\mathcal{B}}_G(\alpha,\theta)$, the feasibility of estimating this variance is still questionable.

We instead suggest a subsampling method to obtain the variance estimate of $\widehat{\mathcal{P}}_G(\alpha,\theta)$ or $\widehat{\mathcal{B}}_G(\alpha,\theta)$. For the notational simplicity, we only consider the variance estimate of the extended bipolarization index $\widehat{\mathcal{B}}_G(\alpha,\theta)$, specifically the jackknife variance estimator. In the literature, it is known that bootstrap variance estimation for the Gini coefficient is still computationally demanding especially when n is large as in the case of conventional income data. This is still the case for $\widehat{\mathcal{P}}_G(\alpha,\theta)$ or $\widehat{\mathcal{B}}_G(\alpha,\theta)$ since we need to calculate the Gini coefficients G and G_k in each iteration step. On the other hand, the jackknife variance estimator can be obtained much faster than the bootstrap variance estimator once we construct an efficient algorithm. In this subsection, we propose a jackknife variance estimation algorithm for the generalized polarization index estimator, but especially for $\widehat{\mathcal{B}}_G(\alpha,\theta)$ for the notational simplicity. The main algorithm for the Gini coefficient part is based on Karagiannis and Kovacevic (2000).

1. Sort the original income data in *ascending* order and denote them as $\{y_i\}_{i=1}^n$; therefore, the index of y_i also represents its rank r_i.

   (a) Calculate the sample mean $\overline{y}=(1/n)\sum_{i=1}^n y_i$.

(b) Define $L = \sum_{i=1}^{n} r_i y_i$ and $H_i = \sum_{j=i+1}^{n} y_j$ for $i = 1, 2, \cdots, n$ with $H_n = 0$.

(c) Then the Gini coefficient can be obtained as $\widehat{G} = (2L) / \left(\overline{y} n^2\right) - (n+1)/n$.

2. Group the data into two (for intervals A_1 and A_2) using the sample mean $\overline{y}$ as the cutoff point (or using any proper method).

   (a) Since the original data is already sorted in step 1, the data is each group is also properly ordered. For each group $k = 1, 2$, we let n_k be the number of observations in group k and $\{y_{k,i}\}_{i=1}^{n_k}$ be the sorted income data in group k. We also denote $r_{k,i}$ as the rank of $y_{k,i}$'s in group k.

   (b) Calculate the group sample proportion $p_k = n_k/n$ and the group sample mean $\overline{y}_k = (1/n_k) \sum_{i=1}^{n_k} y_{k,i}$. Also define $L_k = \sum_{i=1}^{n_k} r_{k,i} y_{k,i}$ and $H_{k,i} = \sum_{j=i+1}^{n_k} y_{k,j}$ for $i = 1, 2, \cdots, n_k$ with $H_{k,n_k} = 0$.

   (c) Then the Gini coefficient of group k can be obtained as $\widehat{G}_k = (2L_k) / \left(\overline{y}_k n_k^2\right) - (n_k + 1)/n_k$.

   (d) Using values obtained in steps 1 and 2, get $\widehat{\mathcal{B}}_G(\alpha, \theta)$ as in (11) for given α and θ.

3. From the entire sample $\{y_t\}_{t=1}^{n}$, omit the i-th observation y_i $(i = 1, 2, \cdots, n)$.

   (a) Using $(n-1)$-number of observations, obtain the new sample mean and the Gini coefficient as

$$\overline{y}_{(-i)} = \frac{1}{n-1} (n\overline{y} - y_i) \quad \text{and} \quad \widehat{G}_{(-i)} = \frac{2}{\overline{y}_{(-i)} (n-1)^2} (L - r_i y_i - H_i) - \frac{n}{n-1}.$$

(b) Let

$$p_{1,(-i)} = \begin{cases} (n_1 - 1) / (n - 1) & \text{if } y_i \in A_1 \text{ (i.e., } y_i < \overline{y}) \\ n_1 / (n - 1) & \text{if } y_i \in A_2 \text{ (i.e., } y_i \geq \overline{y}) \end{cases},$$

$$\overline{y}_{1,(-i)} = \begin{cases} (n_1 \overline{y}_1 - y_i) / (n_1 - 1) & \text{if } y_i \in A_1 \text{ (i.e., } y_i < \overline{y}) \\ \overline{y}_1 & \text{if } y_i \in A_2 \text{ (i.e., } y_i \geq \overline{y}) \end{cases},$$

$$\overline{y}_{2,(-i)} = \begin{cases} \overline{y}_2 & \text{if } y_i \in A_1 \text{ (i.e., } y_i < \overline{y}) \\ (n_2 \overline{y}_2 - y_i) / (n_2 - 1) & \text{if } y_i \in A_2 \text{ (i.e., } y_i \geq \overline{y}) \end{cases}.$$

Then the Gini coefficients of group 1 and 2 can be obtained as

$$\widehat{G}_{1,(-i)} = \begin{cases} \frac{2}{\overline{y}_{1,(-i)}(n_1-1)^2} (L_1 - r_{1,i} y_i - H_{1,i}) - \frac{n_1}{n_1 - 1} & \text{if } y_i \in A_1 \text{ (i.e., } y_i < \overline{y}) \\ \widehat{G}_1 & \text{if } y_i \in A_2 \text{ (i.e., } y_i \geq \overline{y}) \end{cases},$$

$$\widehat{G}_{2,(-i)} = \begin{cases} \widehat{G}_2 & \text{if } y_i \in A_1 \text{ (i.e., } y_i < \overline{y}) \\ \frac{2}{\overline{y}_{2,(-i)}(n_2-1)^2} (L_2 - r_{2,i} y_i - H_{2,i}) - \frac{n_2}{n_2 - 1} & \text{if } y_i \in A_2 \text{ (i.e., } y_i \geq \overline{y}) \end{cases}.$$

(c) Using values obtained in step 3 above, we get $\widehat{\mathcal{B}}_{G,(-i)}(\alpha, \theta)$ as

$$\widehat{\mathcal{B}}_{G,(-i)}(\alpha, \theta) = \left(1 - \frac{\overline{y}_{1,(-i)}}{\overline{y}_{(-i)}}\right) p_{1,(-i)} \left[(1-\theta)\left(\frac{p_{1,(-i)}}{\widehat{G}_{1,(-i)}/\widehat{G}_{(-i)}}\right)^{\alpha} + \theta \left(\frac{1 - p_{1,(-i)}}{\widehat{G}_{2,(-i)}/\widehat{G}_{(-i)}}\right)^{\alpha}\right].$$

4. Iterate step 3 from $i = 1$ to $i = n$ and recursively calculate

$$v_i = v_{i-1} + \frac{n-1}{n}\left(\widehat{\mathcal{B}}_{G,(-i)}(\alpha, \theta) - \widehat{\mathcal{B}}_G(\alpha, \theta)\right)^2$$

with $v_0 = 0$. Then, v_n is the jackknife variance estimator for $\widehat{\mathcal{B}}_c(\alpha, \theta)$.

References

Aghevli, B.B. and F. Mehran (1981). Optimal Grouping of Income Distribution Data, *Journal of the American Statistical Association*, 76, 22-26.

Blau, J.R., and P.M. Blau (1982). The Cost of Inequality: Metropolitan Structure and Violent Crime, *American Sociological Review*, 47, 114-129.

Bourguignon, F., J. Nunez, and F. Sanchez (2003). Which Part of Income Distribution does matter for Explaining Crime? The Case of Columbia, *DELTA Working Paper No. 2003-04*.

Cornwell, C., and W.N. Trumbull (1994). Estimating the Economic Model of Crime with Panel Data, *Review of Economics and Statistics*, 76, 360-366.

Demombynes, G., and B. Ozler (2005). Crime and Local Inequality in South Africa, *Journal of Development Economics*, 76, 265-292.

Doyle, J.M., E.Ahmed, and R.N. Horn (1999). The Effects of Labor Markets and Income Inequality on Crime: Evidence from Panel Data, *Southern Economic Journal*, 65(4), 717-738.

Duclos, J-I., J. Esteban and D. Ray (2004). Polarization: Concepts, Measurement, Estimation, *Econometrica*, 74, 1737-1772.

Ehrlich, I. (1973). Participation in Illegitimate Activities: A Theoretical and Empirical Investigation, *Journal of Political Economy*, 81, 521-65.

Esteban, J. and D. Ray (1994). On the Measurement of Polarization, *Econometrica*, 62, 819-851.

Esteban, J. and D. Ray (1999). Conflict and Polarization, *Journal of Economic Theory*, 87, 379-415.

Esteban, J., C. Gradin and D. Ray (1999). Extensions of a Measure of Polarization with an Application to the Income Distribution of Five OECD Countries, *Working Paper No. 218*, Syracuse University.

Fajnzylber, P., D. Lederman, and N. Loayza (2002). Inequality and Violent Crime, *Journal of Law and Economics*, 45, 1-39.

Foster, J.E., and M.C. Wolfson (1992). Polarization and the Decline of the Middle Class: Canada and the U.S., *mimeo*, Vanderbilt University.

Freeman, R.B. (1996). Why Do So Many Young American Men Commit Crimes and What Might We Do About It?, *Journal of Economic Perspectives*, 10, 25-42.

Gastwirth, J.L., T.K. Nayak and A.M. Kreiger (1986). Large Sample Theory for the Bounds on the Gini and Related Indices of Inequality Estimated from Grouped Data, *Journal of Business and Economic Statistics*, 4, 269-273.

Glaeser, E.L., B. Sacerdote, and J.A. Scheinkman (1996). Crime and Social Interactions, *Quarterly Journal of Economics*, 111(2), 507-548.

Hartigan, J.A. and M.A. Wong (1979). A K-means clustering algorithm, *Applied Statistics*, 28, 100–108.

Karagiannis, E. and M. Kovacevic (2000). A Method to Calculate the Jackknife Variance Estimator for the Gini Coefficient, *Oxford Bulletin of Economics and Statistics*, 62, 119-122.

Kelly, M. (2000). Inequality ad Crime, *Review of Economics and Statistics*, 82, 530-39.

Lee, D.S.-Y. (1993). An Empirical Investigation of the Economic Incentives for Criminal Behavior, *B.A. thesis*, Harvard University.

Nilsson, A. (2004). Income Inequality and Crime: The Case of Sweden, *Institute for Labor Market Policy Evaluation, Working Paper 2004:6.*

Sethuraman, J. (1963). Some Limit Distributions Connected with Fixed Interval Analysis, *Sankhyā*, Ser. A, 27, 395-398.

Wilson, M., and M. Daly (1997). Life Expectancy, Economic Inequality, Homicide, and Reproductive Timing in Chicago Neighborhoods, *British Medical Journal*, 314, 1271.

Wolfson, M.C. (1994). When Inequalities Diverge, *American Economic Review: Papers and Proceedings*, 84, 353-358.

Wolfson, M.C. (1997). Divergent Inequalities: Theory and Empirical Results, *Review of Income and Wealth*, 43, 401-421.