Income Polarization and Crime: A Generalized Index and Evidence from Panel Data*

Yoonseok Lee†　　Donggyun Shin‡

May 2011

Abstract

We develop a generalized polarization index of income distribution to measure social unrest and test two hypotheses: for a given income inequality, polarization captures reduced between-group income mobility raising crime incentives; for a given mobility, the poor feel more alienated against the rich than vice versa. We design the index to allow for asymmetric feeling of alienation and derive its asymptotic properties. Fixed-effects estimates show that the polarization index is significantly correlated with crime rates whereas the Gini index is not; placing more weight on the alienation felt by the lower-income class increases the explanatory power of the new index.

Keywords and phrases: income polarization, crime, asymmetry, social unrest, income mobility.

JEL classification: C51; D63; K14.

*The authors acknowledge valuable comments from Matthew Shapiro, James Hines, Lutz Kilian, Daniel Silverman and Joel Slemrod. The usual disclaimer applies.

†Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109; *email*: `yoolee@umich.edu`

‡Department of Economics, Kyung Hee University, Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Korea; *email*: `dgshin@khu.ac.kr`

1 Introduction

It is commonly thought that income inequality and crime are positively correlated. Standard economic models in the tradition of Becker (1968) imply that, as the income distribution gets more unequal, the gap between the benefits and costs of crime widens and thus the incentive for (property) crime becomes higher. The sociological theory of relative deprivation emphasizes that social tension increases as the poor feel more inferior to the rich, resulting in higher crime rates (e.g., Merton, 1968; Blau and Blau, 1982). Although these theories are compelling, the existing empirical evidence is mixed at best. While most cross-sectional studies conclude that inequality leads to crime (see, e.g., the survey by Demombynes and Özler, 2005), these findings may be subject to an omitted variable bias problem, as they do not control for unobserved heterogeneity that is possibly correlated with inequality. Studies analyze panel data to control for unobserved fixed effects, but often find no significant relationship between income inequality and crime (e.g., Freeman, 1996; Doyle, Ahmed and Horn, 1999; Kelly, 2000; Fajnzylber, Lederman and Loayza, 2002; Demombynes and Özler, 2005).

We argue in this paper that the empirical literature does not generate robust results because common income inequality measures such as the Gini coefficient do not adequately account for the effects of between-group income mobility that individuals experience over their lifetimes.[1] In this paper, we emphasize the bipolarization aspect of income distribution as an important determinant of individual crime behavior. As pointed by Wolfson (1994) and Esteban and Ray (1994), two income distributions with the same level of inequality can have different degrees of bipolarization. Conventional inequality measures describe the overall dispersion of income distribution, and they are used to represent an individual's relative social status as an indicator of unhappiness at a point in time (e.g., Alesina, Di Tella and MacCulloch, 2004). In contrast, bipolarization of income distribution emphasizes within-group clustering as well as the distance between different income groups (e.g., Esteban and Ray, 1994; Esteban, Gradin and Ray, 2007). It thus can capture the phenomenon of the disappearing middle class and formation of two

[1] Examples include the Gini coefficient (e.g., Ehrlich, 1973; Blau and Blau, 1982; Fajnzylber, Lederman, Loayza, 2002), the proportion of the population below a certain percentage of the median income (e.g., Nilsson, 2004; Bourguignon, Nunez, and Sanchez, 2003), and the mean log deviation as a special case of a generalized entropy measure (e.g., Demombynes and Özler, 2005).

segregated income classes, which naturally have a strong implication for between-group income mobility. Even for a given level of overall income inequality, in a society with lower income mobility, the poor have more dismal prospects for the future (i.e., lower expected lifetime income) and feel more deprived than those in a society with higher mobility. In the special case of a society with zero mobility between different income groups, the only way of joining the higher income group is to supply labor in the illegal labor market. Income bipolarization thus properly reflects the level of social unrest, and is expected to explain the crime rate better than simple inequality measures.

However, the income bipolarization index developed by Esteban and Ray (1994) and Esteban, Gradin and Ray (2007) does not explain the crime rate effectively because it does not reflect heterogeneity among different income groups. Specifically, it does not consider asymmetric psychological aspects between the poor and the rich. Degrees of alienation or antagonism are generally different between the low- and the high-income classes, and such difference affects the asymmetry in crime incentives of different income groups. We develop a bipolarization index that properly allows for such asymmetric degrees of alienation so that it will explain the crime rate more effectively.

The main contributions of this paper are three-fold. First, we develop a more general measure of income polarization that accounts for both between-group mobility and asymmetric antagonism between different income groups. It thus measures the level of underlying social unrest more effectively compared to existing inequality or polarization indices, particularly in explaining negative social consequences like crime. Importantly, this new measure nests existing measures as special cases (e.g., Gini coefficient and polarization index by Esteban and Ray, 1994). Second, we derive the asymptotic distribution of the new measure so that standard statistical inferences can be made, and develop an easy-to-implement variance estimation algorithm. Despite the repeated reports that income distribution has become more bipolarized (e.g., Esteban, Gradin and Ray, 2007), few empirical studies provide formal statistical conclusions due to the fact that the indices they examine lack theoretical distributions. The asymptotic result is readily applied to existing indices because the new index nests existing measures as special cases. Third, we empirically verify the relationship between this new measure and crime rates. Analysis based on panel data sets at the country level and the U.S. state level reveals that when both the Gini and the bipolarization indices are included in the regression model, bipolarization significantly raises the crime rate whereas the Gini index does not. We also find that our generalized bipolarization index has more explanatory power when a heavier weight is placed on the feeling of alienation of the lower income class.

The remainder of this paper is organized as follows. Section 2 develops a new bipolarization index to represent the level of social unrest or aggregated effective antagonism in a society. The asymptotic distribution of the index and an easy-to-implement jackknife variance estimation algorithm are also obtained. A description of the data, the estimation method, and empirical findings are presented in Section 3. Section 4 concludes the paper by summarizing the main results.

2 Measuring Social Unrest: Generalized Polarization Index

2.1 Generalized Polarization Index

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 $2 \leq K \leq n$. 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 and the group mean μ_k as

$$\pi_k = \int_{a_{k-1}}^{a_k} dF(y) \quad \text{and} \quad \mu_k = \frac{1}{\pi_k} \int_{a_{k-1}}^{a_k} y dF(y),$$

where we assume that $\pi_k > 0$ for all k. It follows that $\sum_{k=1}^K \pi_k = 1$ and the overall mean is defined as $\mu = \sum_{k=1}^K \pi_k \mu_k = \int y dF(y)$. Note that the group means are in ascending order by construction so that $\mu_k < \mu_j$ if $k < j$.

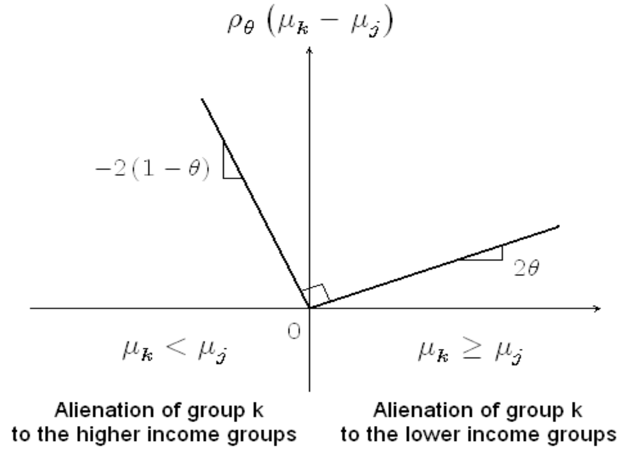

Figure 1: $\rho_\theta\left(\mu_k-\mu_j\right)$ describes asymmetric alienations of group k

We define the polarization index $\mathcal{P}$ as

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

where $0\leq\theta\leq1/2$ and $\alpha\geq0$ are parameters chosen by a researcher. Here $\rho_\theta(u)=2u(\theta-\mathbb{I}\{u<0\})$ represents the between-group alienation measure, which depends on the income distance between different income groups. $\mathbb{I}\{\cdot\}$ is the binary indicator. $\phi_k>0$ represents the (relative-) identity measure of group k, where α is the weight placed on the identity measure relative to the alienation measure. Similar to Esteban and Ray (1994), therefore, $\mathcal{P}(\alpha,\theta)$ combines the following two concepts: within-group *identity* and between-group *alienation*. Note that if $\theta=1/2$ (and thus $\rho_\theta(u)=|u|$) and $\phi_k=\pi_k$, then the polarization index $\mathcal{P}(\alpha,\theta)$ in (1) is the same as the index developed by Esteban and Ray (1994), $ER_K(\alpha)=(1/\mu)\sum_{k=1}^{K}\sum_{j=1}^{K}\pi_k^{1+\alpha}\pi_j\left|\mu_k-\mu_j\right|$. Furthermore, if $\theta=1/2$ and $\alpha=0$, then the index becomes the Gini index for grouped data.[2]

[2] It is worth noting that $\mathcal{P}(\alpha,\theta)$ satisfies the axioms of polarization by Esteban and Ray (1994), provided $0<\alpha\leq1.6$. It is because $\mathcal{P}(\alpha,\theta)$ conforms to the general form of the index satisfying the axioms, $C\sum_{k=1}^{K}\sum_{j=1}^{K}\pi_k\pi_j\mathcal{I}^\alpha(\pi_k)\mathcal{A}\left(\mu_k,\mu_j\right)$ for some constant C, if the identity function is defined as $\mathcal{I}(\pi_k)=\phi_k$ and the alienation function as $\mathcal{A}\left(\mu_k,\mu_j\right)=\rho_\theta\left(\mu_k-\mu_j\right)$.

To be more specific, we formulate between-group alienation as

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

which is more general than $ER_K(\alpha)$ in the sense that we allow for *asymmetric* feelings of alienation, with the degree of the asymmetry being determined by the value θ. Specifically, since we let $0\leq\theta\leq 1/2$, the lower income groups feel more alienated from the higher income groups than vice versa. The asymmetry gets more severe as θ goes to zero; a lower θ can be understood as a relatively higher level of antagonism of the low income group compared to that of the high income group. As an extreme case, if $\theta=0$ then the richer groups do not feel any alienation against the poorer groups. If $\theta=1/2$ then the degree of alienation is symmetric between the groups, which corresponds to the existing polarization indices (e.g., Esteban and Ray, 1994) and income inequality measures. Different from the existing income polarization indices, therefore, the polarization index $\mathcal{P}(\alpha,\theta)$ reflects not only the between-group income distance but also the asymmetry in the antagonism each group has against the others. Figure 1 depicts $\rho_\theta(\mu_k-\mu_j)$, where the absolute value of the slope determines the degree of the asymmetric alienation of group k to different income-level groups. Note that the parameter of asymmetric feeling of alienation, θ, is different from the inequality aversion parameter in Atkinson's index (Atkinson, 1970) or generalized entropy index (Cowell and Kuga, 1981; Shorrocks, 1984); the latter measures the overall (and thus symmetric) inequality aversion level whereas the former measures the asymmetric inequality aversion levels in each direction.

For the within-group identity parameter ϕ_k, we assume that the degree of group-identity is positively affected by the group size as in the Esteban-Ray type indices but is inversely related to within-group relative income dispersion. More precisely, we let

$$\phi_k=\frac{\pi_k}{(G_k/G)}, \tag{3}$$

where G is the Gini index of the entire population and G_k is the Gini index of the income distribution over the interval A_k. In this specification, within-group identity gets larger either when the population share of group k increases or when the relative dispersion of the within-group income distribution of group k decreases. The population proportion π_k reflects the majority or minority of the group in the society. An individual feels the income class separation more as a social structural problem when the population proportion of the income group that she belongs to gets larger. The relative dispersion of group k, G_k/G, represents the degree of clustering of individuals' income levels in group k around its group mean. An individual identifies more with her group members as within-group income levels become more similar. Note that the *relative* dispersion measure (G_k/G) is more meaningful than the *absolute* dispersion measure (G_k) since changes in G_k also alter the overall income inequality level G.

The forms of between-group alienation $(\rho_\theta\left(\mu_k - \mu_j\right))$ and within-group identity (ϕ_k) support the idea that the generalized polarization index $\mathcal{P}\left(\alpha, \theta\right)$ represents the level of social unrest (or the level of total antagonism) implied by income distribution more effectively than existing indices, such as Esteban and Ray (1994) or Esteban, Gradin, and Ray (2007). Specifically, $\mathcal{P}\left(\alpha, \theta\right)$ allows not only for asymmetric degrees of alienation among different income groups but also for a more plausible identification function.

From (2) and (3), we can readily obtain an estimator for $\mathcal{P}\left(\alpha, \theta\right)$ in (1) using proper estimators for π_k, μ_k, G_k $(k = 1, 2, \cdots, K)$ and G as

$$\widehat{\mathcal{P}}_n\left(\alpha, \theta\right) = \frac{1}{2\overline{y}} \sum_{k=1}^{K} \sum_{j=1}^{K} \widehat{\pi}_k \widehat{\pi}_j \left(\frac{\widehat{\pi}_k}{\widehat{G}_k / \widehat{G}} \right)^{\alpha} \rho_\theta \left(\overline{y}_k - \overline{y}_j \right), \tag{4}$$

where

$$\widehat{\pi}_k = \frac{n_k}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{ y_i \in A_k \right\}, \quad \overline{y}_k = \frac{1}{n_k} \sum_{i=1}^{n} y_i \mathbb{I}\left\{ y_i \in A_k \right\}$$

and $\overline{y} = (1/n) \sum_{i=1}^{n} y_i$ with n_k being the number of observations in the interval A_k. Recall that $\widehat{G}_k = \left(1/(2\overline{y}_k n_k^2)\right) \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| \mathbb{I}\left\{y_i \in A_k\right\} \mathbb{I}\left\{y_j \in A_k\right\}$ is the Gini coefficient for group k and $\widehat{G} = \left(1/(2\overline{y} n^2)\right) \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$ is the standard Gini coefficient for the entire income distribution. Since K is assumed to be fixed and independent of n, we can assume that $n_k \to \infty$ for all $k = 1, 2, \cdots, K$ as $n \to \infty$ without loss of generality. Therefore, for each given

(α, θ), the consistency of $\widehat{\mathcal{P}}_n(\alpha, \theta)$ for the true polarization index $\mathcal{P}(\alpha, \theta)$ readily follows by applying the continuous mapping theorem, since all the individual estimators are consistent.

Remark 1 In the polarization index $ER_K(\alpha)$ by Esteban and Ray (1994), π_k measures the feeling of identification each individual has toward her own group members. In comparison, our new polarization index $\mathcal{P}(\alpha, \theta)$ considers not only the 'size effect' (π_k) but also the relative 'clustering effect' (G_k/G) in measuring the degree of within-group identity. In this regards, it can be understood that the extended polarization index by Esteban, Gradin and Ray (2007) also considers the group clustering effect implicitly, though such a point is not discussed in their paper. Note that their extended index is defined as $EGR_K(\alpha, \beta) = ER_K(\alpha) - \beta C_K$, where $\beta(> 0)$ is some arbitrary weight parameter (often $\beta = 1$) and C_K is the error in approximating the continuous Lorenz curve by K-piecewise linear functions. C_K gets smaller as within-group income distributions become more clustered around their group means. Some distinctions, however, are noticed between $\mathcal{P}(\alpha, \theta)$ and $EGR_K(\alpha, \beta)$. First, $\mathcal{P}(\alpha, \theta)$ considers relative clustering whereas $EGR_K(\alpha, \beta)$ considers absolute clustering. Second, $\mathcal{P}(\alpha, \theta)$ constructs the identity function of each group using its own clustering effect whereas $EGR_K(\alpha, \beta)$ combines all the clustering effects to form the overall approximation error. Third, it is known that the arbitrariness of an additional parameter β could generate some undesirable properties. For example, $EGR_K(\alpha, \beta)$ could be negative or it could violate one of the axioms in Esteban and Ray (1994). The index could even decrease when the between-group distance increases (e.g., Lasso de la Vega and Urrutia, 2006). In contrast, $\mathcal{P}(\alpha, \theta)$ satisfies all the axioms in Esteban and Ray (1994) since the structure of the index remains the same as also noted in footnote 2.

Remark 2 To estimate the polarization index, it is required that the intervals $A_k = [a_{k-1}, a_k)$ are pre-determined. Even when the number of intervals K is chosen, finding the cutoff points $\{a_k\}_{k=1}^{K-1}$ is still a difficult problem (e.g., K-means clustering algorithm; Hartigan and Wong, 1979). To solve this problem, Esteban, Gradin and Ray (2007) employ Aghevli and Mehran (1981)'s method of optimal grouping for a given K. The idea is that one minimizes the sum of within-group income dispersions (e.g., the mean difference) 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 that minimize the overall approximation error. When we consider the bipolarization index, in particular, the number of intervals is fixed as $K = 2$ and we only need to choose one cutoff point $a_1^* (= y^*)$, which separates the entire income distribution into two. Aghevli and Mehran (1981) show that the optimal cutoff point is the population mean (i.e., $y^* = \mu$) in this case.

2.2 Asymptotic Distribution of Bipolarization Index

Despite the repeated reports that income distribution has become more bipolarized, few studies provide formal statistical conclusions due to the fact that their indices lack theoretical distributions. This subsection derives the asymptotic distribution of the generalized index estimator $\widehat{\mathcal{P}}_n(\alpha, \theta)$ to facilitate further statistical inferences for the index. Note that this new index is general enough to include the existing Esteban-Ray type indices as special cases. The statistical results below can thus be directly applied to those indices. In what follows, we focus on the income bipolarization, which considers the case of $K = 2$. The results can be readily generalized to the cases of $K > 2$. More precisely, for a given threshold y^*, the bipolarization index is defined as

$$\begin{aligned}\mathcal{B}(\alpha, \theta) &= \frac{1}{2\mu} \sum_{k=1}^{2} \sum_{j=1}^{2} \pi_k \pi_j \left(\frac{\pi_k}{G_k/G} \right)^{\alpha} \rho_\theta (\mu_k - \mu_j) \\ &= \frac{\mu_2 - \mu_1}{\mu} \pi_1 \pi_2 \left[(1-\theta) \left(\frac{\pi_1}{G_1/G} \right)^{\alpha} + \theta \left(\frac{\pi_2}{G_2/G} \right)^{\alpha} \right]\end{aligned}$$

for $\mu_1 < \mu_2$ by construction. Similarly as $\widehat{\mathcal{P}}_n(\alpha, \theta)$, its consistent estimator can be obtained as

$$\begin{aligned}\widehat{\mathcal{B}}_n(\alpha, \theta) &= \frac{\overline{y}_2 - \overline{y}_1}{\overline{y}} \widehat{\pi}_1 \widehat{\pi}_2 \left[(1-\theta) \left(\frac{\widehat{\pi}_1}{\widehat{G}_1/\widehat{G}} \right)^{\alpha} + \theta \left(\frac{\widehat{\pi}_2}{\widehat{G}_2/\widehat{G}} \right)^{\alpha} \right] \\ &= \left(1 - \frac{\overline{y}_1}{\overline{y}} \right) \widehat{\pi}_1 \widehat{G} \left[(1-\theta) \left(\frac{\widehat{\pi}_1}{\widehat{G}_1} \right)^{\alpha} + \theta \left(\frac{1 - \widehat{\pi}_1}{\widehat{G}_2} \right)^{\alpha} \right].\end{aligned} \tag{5}$$

We let $\widehat{g} = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$ be the standard mean difference coefficient and $\widehat{g}_{k\ell} = \widehat{g}_{\ell k} = n_k^{-1} n_\ell^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| \mathbb{I}\{y_i \in A_k\} \mathbb{I}\{y_i \in A_\ell\}$ be the sub-group mean difference coefficient for all $k, \ell = 1, \cdots, K$. It holds that $\widehat{g} = \sum_{k=1}^{K} \sum_{\ell=1}^{K} \widehat{g}_{k\ell} \widehat{\pi}_k \widehat{\pi}_\ell$ (e.g., Dagum, 1997) and $\overline{y} = \sum_{k=1}^{K} \widehat{\pi}_k \overline{y}_k$. Then, $\widehat{\mathcal{B}}_n(\alpha, \theta)$ in (5) can be rewritten as

$$\widehat{\mathcal{B}}_n(\alpha, \theta) = \frac{(\overline{y}_2 - \overline{y}_1)\,\widehat{\pi}_1 \widehat{\pi}_2}{(\widehat{\pi}_1 \overline{y}_1 + \widehat{\pi}_2 \overline{y}_2)^{1+\alpha}} \left(\widehat{\pi}_1^2 \widehat{g}_{11} + 2\widehat{\pi}_1 \widehat{\pi}_2 \widehat{g}_{12} + \widehat{\pi}_2^2 \widehat{g}_{22}\right)^{\alpha} \left\{ (1-\theta) \left(\frac{\widehat{\pi}_1 \overline{y}_1}{\widehat{g}_{11}} \right)^{\alpha} + \theta \left(\frac{\widehat{\pi}_2 \overline{y}_2}{\widehat{g}_{22}} \right)^{\alpha} \right\}$$

since $\widehat{G} = \widehat{g}/2\overline{y}$ and $\widehat{G}_k = \widehat{g}_{kk}/2\overline{y}_k$ for each $k = 1, 2$. In order to derive the asymptotic distribution of $\widehat{\mathcal{B}}_n(\alpha, \theta)$, we introduce the following U-statistics for some given cutoff point y^*:

$$\begin{aligned}
U_1 &= n^{-1} \textstyle\sum_{i=1}^{n} \mathbb{I}\{y_i \le y^*\} \\
U_2 &= n^{-1} \textstyle\sum_{i=1}^{n} y_i \mathbb{I}\{y_i \le y^*\} \\
U_3 &= n^{-2} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|\, \mathbb{I}\{y_i \le y^*\}\, \mathbb{I}\{y_j \le y^*\} \\
U_4 &= n^{-1} \textstyle\sum_{i=1}^{n} \mathbb{I}\{y_i > y^*\} \\
U_5 &= n^{-1} \textstyle\sum_{i=1}^{n} y_i \mathbb{I}\{y_i > y^*\} \\
U_6 &= n^{-2} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|\, \mathbb{I}\{y_i > y^*\}\, \mathbb{I}\{y_j > y^*\} \\
U_7 &= n^{-2} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|\, \mathbb{I}\{y_i > y^*\}\, \mathbb{I}\{y_j \le y^*\},
\end{aligned}$$

where denoting

$$\begin{aligned}
\upsilon_1 &= \textstyle\int_{-\infty}^{y^*} dF(y) = F(y^*) \\
\upsilon_2 &= \textstyle\int_{-\infty}^{y^*} y\,dF(y) \\
\upsilon_3 &= \textstyle\int_{-\infty}^{y^*} \int_{-\infty}^{y^*} |y - x|\, dF(y)\, dF(x) \\
\upsilon_4 &= \textstyle\int_{y^*}^{\infty} dF(y) = 1 - F(y^*) \\
\upsilon_5 &= \textstyle\int_{y^*}^{\infty} y\,dF(y) \\
\upsilon_6 &= \textstyle\int_{y^*}^{\infty} \int_{y^*}^{\infty} |y - x|\, dF(y)\, dF(x) \\
\upsilon_7 &= \textstyle\int_{-\infty}^{y^*} \int_{y^*}^{\infty} |y - x|\, dF(y)\, dF(x),
\end{aligned}$$

$U_1, U_2, \cdots, U_7$ are consistent estimators of $\upsilon_1, \upsilon_2, \cdots, \upsilon_7$, respectively. Theorem 7.1 of Hoeffding (1948) gives joint asymptotic distribution of $U_1, U_2, \cdots, U_7$ as follows.

Lemma 1 *Let $\{y_i\}_{i=1}^{n}$ be i.i.d. with continuous distribution $F(y)$ and finite variance. If $n_1, n_2 \to \infty$ as $n \to \infty$ and $\epsilon < F(y^*) < 1 - \epsilon$ for some $\epsilon \in (0, 1)$, then the joint distribution of $\sqrt{n}(U_m - \upsilon_m)$ for $m = 1, 2, \cdots, 7$ tends to the 7-variate normal distribution as $n \to \infty$ with* *zero mean and covariance matrix* Σ *given by*

$$\begin{pmatrix} \upsilon_1(1-\upsilon_1) & \upsilon_2(1-\upsilon_1) & 2\upsilon_3(1-\upsilon_1) & -\upsilon_1\upsilon_4 & -\upsilon_1\upsilon_5 & -\upsilon_1\upsilon_6 & -\upsilon_1\upsilon_7 \\ & \xi_2 & 2\xi_{2,3} & -\upsilon_2\upsilon_4 & -\upsilon_2\upsilon_5 & -\upsilon_2\upsilon_6 & -\upsilon_2\upsilon_7 \\ & & 4\xi_3 & -\upsilon_3\upsilon_4 & -\upsilon_3\upsilon_5 & -\upsilon_3\upsilon_6 & -\upsilon_3\upsilon_7 \\ & & & \upsilon_4(1-\upsilon_4) & \upsilon_5(1-\upsilon_4) & 2\upsilon_6(1-\upsilon_4) & \upsilon_7(1-\upsilon_4) \\ & & & & \xi_5 & 2\xi_{5,6} & 2\upsilon_7(1-\upsilon_5) \\ & & & & & 4\xi_6 & 4\xi_{6,7} \\ & & & & & & 4\xi_7 \end{pmatrix},$$

where

$$\begin{aligned}
\xi_2 &= \textstyle\int_{-\infty}^{y^*} y^2 dF(y) - \upsilon_2^2 \\
\xi_3 &= \textstyle\int_{-\infty}^{y^*} \left\{\int_{-\infty}^{y^*} |y-x|\, dF(x)\right\}^2 dF(y) - \upsilon_3^2 \\
\xi_{2,3} &= \textstyle\int_{-\infty}^{y^*}\int_{-\infty}^{y^*} y\,|y-x|\, dF(y)\, dF(x) - \upsilon_2\upsilon_3 \\
\xi_5 &= \textstyle\int_{y^*}^{\infty} y^2 dF(y) - \upsilon_5^2 \\
\xi_6 &= \textstyle\int_{y^*}^{\infty} \left\{\int_{y^*}^{\infty} |y-x|\, dF(x)\right\}^2 dF(y) - \upsilon_6^2 \\
\xi_7 &= \textstyle\int_{y^*}^{\infty} \left\{\int_{-\infty}^{y^*} (y-x) dF(x)\right\}^2 dF(y) - \upsilon_7^2 \\
\xi_{5,6} &= \textstyle\int_{y^*}^{\infty}\int_{y^*}^{\infty} y\,|y-x|\, dF(y)\, dF(x) - \upsilon_5\upsilon_6 \\
\xi_{6,7} &= \textstyle\int_{y^*}^{\infty} \left\{\int_{y^*}^{\infty} |y-x|\, dF(y)\right\}\left\{\int_{-\infty}^{y^*} (y-x) dF(y)\right\} dF(x) - \upsilon_6\upsilon_7.
\end{aligned}$$

Note that Bishop, Formby and Zheng (1997) consider the joint distribution of U_1, U_2, U_3; Lemma 1 extends their result.

Since $\widehat{\pi}_1 = U_1$, $\widehat{\pi}_2 = U_4$, $\overline{y}_1 = U_2/U_1$, $\overline{y}_2 = U_5/U_4$, $\widehat{g}_{11} = U_3/U_1^2$, $\widehat{g}_{12} = U_7/U_1U_4$ and $\widehat{g}_{22} = U_6/U_4^2$, we can rewrite $\widehat{\mathcal{B}}_n(\alpha,\theta)$ as

$$\widehat{\mathcal{B}}_n(\alpha,\theta) = \frac{U_5U_1 - U_2U_4}{(U_2+U_5)^{1+\alpha}}(U_3+2U_7+U_6)^{\alpha}\left\{(1-\theta)\left(\frac{U_2U_1^2}{U_3}\right)^{\alpha} + \theta\left(\frac{U_5U_4^2}{U_6}\right)^{\alpha}\right\},$$

which is consistent for (provided $\upsilon_1, \cdots, \upsilon_7 > 0$)

$$\mathcal{B}(\alpha,\theta) = \frac{\upsilon_5\upsilon_1 - \upsilon_2\upsilon_4}{(\upsilon_2+\upsilon_5)^{1+\alpha}}(\upsilon_3+2\upsilon_7+\upsilon_6)^{\alpha}\left\{(1-\theta)\left(\frac{\upsilon_2\upsilon_1^2}{\upsilon_3}\right)^{\alpha} + \theta\left(\frac{\upsilon_5\upsilon_4^2}{\upsilon_6}\right)^{\alpha}\right\} \tag{6}$$

from Lemma 1. Using Lemma 1 above and Theorem 7.5 of Hoeffding (1948), we obtain asymptotic distribution of $\widehat{\mathcal{B}}_n(\alpha,\theta)$ as follows.

Theorem 2 *Under the same conditions as Lemma 1, we have* $\sqrt{n}(\widehat{\mathcal{B}}_n(\alpha,\theta) - \mathcal{B}(\alpha,\theta)) \to_d \mathcal{N}(0, V_{\mathcal{B}}(\alpha,\theta))$ *as* $n \to \infty$*, where* $V_{\mathcal{B}}(\alpha,\theta) = [\nabla\mathcal{B}(\alpha,\theta)]'\Sigma[\nabla\mathcal{B}(\alpha,\theta)]$ *and the* 7×1 *vector* $\nabla\mathcal{B}(\alpha,\theta)$ *is given by*

$$\frac{\zeta g^{\alpha}}{\mu^{1+\alpha}}\begin{pmatrix} \upsilon_5/\zeta + 2\alpha/\upsilon_1 & \upsilon_5/\zeta \\ -\upsilon_4/\zeta - (1+\alpha)/\mu + \alpha/\upsilon_2 & -\upsilon_4/\zeta - (1+\alpha)/\mu \\ \alpha/\zeta - \alpha/\upsilon_3 & \alpha/\zeta \\ -\upsilon_2/\zeta & -\upsilon_2/\zeta + 2\alpha/\upsilon_4 \\ \upsilon_1/\zeta - (1+\alpha)/\mu & \upsilon_1/\zeta - (1+\alpha)/\mu + \alpha/\upsilon_5 \\ \alpha/\zeta & \alpha/\zeta - \alpha/\upsilon_6 \\ 2\alpha/\zeta & 2\alpha/\zeta \end{pmatrix}\begin{pmatrix} (1-\theta)\left(\upsilon_2\upsilon_1^2/\upsilon_3\right)^{\alpha} \\ \theta\left(\upsilon_5\upsilon_4^2/\upsilon_6\right)^{\alpha} \end{pmatrix}$$

with $\zeta = (\mu_2 - \mu_1)\pi_1\pi_2$.

Proof. From (6), the partial derivative vector is obtained as $\nabla\mathcal{B}(\alpha,\theta)$, where $\upsilon_3+2\upsilon_7+\upsilon_6 = g$ and $\upsilon_2+\upsilon_5 = \mu$ by construction and by letting $\zeta \equiv \upsilon_5\upsilon_1 - \upsilon_2\upsilon_4 = (\mu_2-\mu_1)\pi_1\pi_2$. Then the result follows immediately from Lemma 1 above using Theorem 7.5 of Hoeffding (1948). Note that $\upsilon_1, \upsilon_4 > 0$ since it is assumed that $\epsilon < F(y^*) < 1-\epsilon$ for some $\epsilon \in (0,1)$, and thus $\upsilon_m > 0$ for $m = 1, 2, \cdots, 7$. ■

In some cases, we need to obtain the standard error of the bipolarization index. For example, when we want to compare bipolarization indices between two different groups or test changes in bipolarization over time, the standard error is a key ingredient for constructing any test statistics. The asymptotic variance $V_{\mathcal{B}}(\alpha,\theta)$ can be consistently estimated using the sample counterparts of υ_m's for $m = 1, 2, \cdots, 7$ (i.e., their U-statistics, $U_1, U_2, \cdots, U_7$), but the calculation is very complicated as appears in Theorem 2. To facilitate the variance estimation of $\widehat{\mathcal{B}}_n(\alpha,\theta)$, we propose a subsampling method, specifically the jackknife variance estimation. The procedure is summarized in detail in the Appendix.

Some remarks are in order. The Esteban and Ray (1994) index $ER_K(\alpha)$ is a special case of $\mathcal{B}(\alpha, \theta)$ when $K = 2$ and $\theta = 1/2$. Therefore, the asymptotic distribution in Theorem 2 and the jackknife variance estimation procedure can be easily extended to $ER_K(\alpha)$. Also note that, while we consider the case with fixed K but large n asymptotics, Duclos, Esteban and Ray (2004) derive the asymptotic properties of $ER_K(\alpha)$ under the large K asymptotics.

3 Empirical Evidence

This section empirically verifies that the new bipolarization measure describes well the relationship with the crime rate that is predicted by theory. More precisely, the goal of the data analysis is to test two hypotheses: crime rates are better explained by bipolarization than inequality; low income earners feel more alienated from high income earners than the latter do from the former, so that placing a heavier weight on lower income earners better explains the crime rate.

To support the first hypothesis, which reflects the economic aspect of income mobility, we need to show that the new bipolarization index is significantly associated with crime rates, controlling for the Gini index and other variables. Based on the Becker model, existing studies (e.g., Imrohoroglu, Merlo and Rupert, 2000; Kelly, 2000; Fajzylber, Lederman and Loayza, 2002) often interpret income inequality as the difference between the gains from crime and its opportunity costs (i.e., the net gain from crime). We claim, however, that the net (expected) gain from crime is better represented by bipolarization rather than inequality. This is so because the degree of between-group income mobility, which is not fully captured by between-group income distance itself, is the key aspect in understanding the net gain.

To test the second hypothesis, which reflects the sociological theory of relative deprivation, we estimate the model for various values of θ and show that the new bipolarization index has more explanatory power as a heavier weight is placed on the poor. However, the rich cannot be entirely neglected in explaining the crime rate and we expect that there exists an optimal weight on the poor (i.e., θ between 0 and 0.5) that best explains the crime rate.

3.1 Data and Econometric Model

The most challenging aspect of this empirical study is to compute the new bipolarization index for each unit of analysis in which crime is examined. In particular, the index requires dividing the sample into the poor and the rich in each unit, and computing various group-level statistics including the Gini coefficient for each group. For this reason, country level data can be very useful because the sample size of micro data on individual households is large in general.

On the contrary, differences in the definition of household income often make it difficult to compare various inequality and polarization measures across different countries. For that reason, we examine country data from the Luxembourg Income Study (LIS) database, which includes income micro data from a large number of countries at multiple points in time using the same definition and components of household income across countries. We use household disposable income adjusted by the 'OECD equivalency scale' (see, e.g., Atkinson, Rainwater and Smeeding, 1995) and household sample weights.

There is one more advantage of using country data. When the unit of analysis is very small, the local crime rate does not necessarily reflect the region's economic conditions. For example, criminals travel to neighborhoods in search of higher returns (e.g., Demombynes and Özler, 2005); those who are frustrated in one region move on to another region, where they have better prospects and thus decide to supply labor to the legal labor market. It generates geographical interdependence (e.g., Glaeser, Sacerdote and Scheinkman, 1996) that makes the analysis extremely complicated. However, the crime market can reasonably be assumed to be closed at the country level in general.

Because our primary goal is not to distinguish empirically between the economic and sociological explanations of the relationship between crime and income distribution, and because the new bipolarization measure reflects both economic and psychological motives of crime, the overall crime rate is the relevant variable to be used for the test of hypotheses. Following Fajzylber, Lederman, and Loayza (2002), however, we primarily analyze violent crime. This is because property crime categories such as theft and burglary suffer from severe underreporting and their definitions or classifications vary depending on the reliability of each country's judicial system.

To further reduce the potential bias in the results that may arise from using country data, we select 13 countries in the LIS (Australia, Austria, Canada, Denmark, France, Finland, Germany, Netherlands, Norway, Sweden, Switzerland, United Kingdom, and United States) that share relatively similar legal traditions and cultural backgrounds, which allows us to focus on aspects of income distribution as a potential crime determinant. The final sample includes 68 country-year observations for 13 countries, where the sampling frequency is five years on average and both the 1980s and the 1990s are covered for most countries.

The crime data we use are from the United Nations Surveys on Crime Trends and the Operations of Criminal Justice Systems (CTS). Regarding other control variables, police size (*POLICE*) is imported from the CTS; average years of education (*EDUCATION*), percentage of urban population among the total population (*URBAN POPULATION*), and population density (*DENSITY*) come from the World Bank; the proportion of men aged between 15 and 29 among the total population (*YOUNG MEN*), unemployment rates (*UNEMPLOYMENT*), and the percentage of tax revenue among gross domestic product (*TAX-GDP-RATIO*), to be used an instrumental variable for *POLICE*, are drawn from the Organization for Economic Cooperation and Development (OECD).

[Figure 2 is about here]

Figure 2 displays estimated Gini and bipolarization indices for selected countries. Each country contains four series of indices, Gini and three bipolarization indices for different values of $\theta = 0.5$, 0.25 and 0. α is set to be 1.6. It reveals that Gini and bipolarization indices have different patterns, although they are related. Focusing on temporal patterns, estimated Gini and bipolarization indices often move in opposite directions. Moreover, three series of bipolarization indices for different θ values show different temporal patterns for most countries.

Although the main analysis is based on country data, we also look for evidence from U.S. state-level data to supplement the main results. The Current Population Survey (CPS) data are used to calculate our bipolarization indices by state and year. Excluding 11 states, which do not have valid observations for some relevant control variables, we work with annual observations on 40 states from 1991 through 2005. The first year the CPS reports state codes is 1991. Compared with the country level data, U.S. data have the advantage of fairly consistent definitions of crime categories and similar crime-related legal traditions across different states. Crime markets are relatively closed at the state level, although less so than at the country level. As the CPS does not report disposable income, before-tax-transfer income is used, although the former is more relevant for crime behavior. The final sample consists of 600 state-year balanced observations for 40 states. The crime data, along with police size, are directly drawn from the Federal Bureau of Investigation Uniform Crime Reports for the period of 1991 through 2005. Among other control variables, *DENSITY*, *YOUNG MEN*, and the percentage of those who have earned bachelor's degrees or higher among the population over 25 years old, all measured by state and by year, are drawn from the U.S. Census Bureau; and state unemployment rates are from the Bureau of Labor Statistics.

For the regression equation, the logarithm of the crime rate is explained by the bipolarization index as well as other demographic, economic and crime-prevention variables. As one of the primary goals is to test if crime is better explained by bipolarization than inequality, the final model includes the Gini index as an additional explanatory variable. In the estimation, the simultaneity between the crime rate and the crime deterrent variable, police size, is addressed using instrumental variables described below. Unobservable unit-specific fixed effects are included in every regression model and 2-step fixed-effects Generalized Method of Moments (GMM) estimation is applied with Heteroskedasticity and Autocorrelation Consistent (HAC) standard error estimates.

3.2 Findings

Estimates in Table 1 through 3 are based on country data, while those in Table 4 are obtained from the U.S. state-level data. For all tables, we let $\alpha = 1.6$ and $y^* = \overline{y}$ in computing the bipolarization index. The first three columns of Table 1 report the base model, where only the conventional Gini and the new bipolarization indices are included to explain the crime rate. Overall, the bipolarization index explains the crime rate significantly, while the Gini index does not. In addition, the explanatory power of the bipolarization index is higher when a greater weight is placed on the poor than an equal weight ($\theta = 0.5$). Compared with the case of $\theta = 0.25$, however, the explanatory power of the bipolarization index is somewhat reduced when the rich are entirely eliminated in designing our bipolarization index ($\theta = 0$).

[Table 1 is about here]

In the last three columns of Table 1, we consider five additional control variables: unemployment rates, population density, the proportion of the urban population among the total, the proportion of young men aged between 15 and 29, and the average years of schooling. While the estimated signs accord with theoretical predictions, the estimated coefficients on these variables are relatively imprecise. All of the results obtained from the basic model are still preserved even in the extended model.

Existing studies, on U.S. data in particular, often emphasize the importance of police activity as a crime determinant, focusing on the effectiveness of crime deterrent efforts measured by police size, the arrest rate, the imprisonment rate, and so on. Omitting these effects may bias the estimated coefficients of our bipolarization and the Gini index if the demand for public safety is correlated with these variables. The only police-related variable available at the country level is the police size, as measured by the number of policemen per 100,000 people. Table 2 reports the results with, in addition to the five control variables used in Table 1, the police size as an additional determinant.

[Table 2 is about here]

Although, for brevity, we report the estimated effects of only the new bipolarization and the Gini index along with police size, little change occurs on the estimated coefficients of other control variables in Table 1. In the first three rows, we treat police size as an exogenous variable. Estimates in the remaining table are obtained by applying the fixed-effects GMM with *TAX-GDP-RATIO* as an instrumental variable (IV) for police size. As argued by Cornwell and Trumbull (1994), countries with residents who have greater preferences for law enforcement will reveal their preferences by voting for higher tax rates to finance larger police forces.

Such countries would have larger police sizes for reasons not directly related to the crime rate. Including police size as an additional exogenous control variable does not change the estimated coefficients on our bipolarization index and Gini index obtained in Table 1, and the coefficient on police size appears positive. In the fixed-effects GMM estimates on the last three rows, however, the estimated effect of police size is now negative, though not significant. Most importantly, all our previous findings in Table 1 are preserved, whether or not police size is treated endogenously.

Table 3 shows the results for total crime. Using total crime as a new dependent variable allows us to use an additional IV for the police variable, *CRIME COMPOSITION*, which is also suggested by Cornwell and Trumbull (1994). Crime composition is defined by the ratio of the number of violent crimes to that of property crimes. While this ratio is not directly related with the total number of crimes, crime composition is believed to be related with police size. For example, with the total number of crimes fixed, a greater proportion of violent crimes calls for more police activity and for more policemen involved.

[Table 3 is about here]

Due to the noise associated with property crime data previously mentioned, the estimated effects of the bipolarization index are reduced by a large amount, but again all of our previous results still hold: it is bipolarization and not inequality that matters in explaining crime. This is more apparent when a greater weight is put on the poor and, for a given level of mobility, even high income earners contribute to the positive effects of bipolarization on crime to some degree.

[Table 4 is about here]

Finally, Table 4 shows evidence from the U.S. data. For brevity, we report the results only for the overall crime rate and the estimated coefficients on the two distribution-related indices along with police size. Table 4 adopts a similar specification as in Table 3. In this case, we use the crime composition of a state and a population-weighted average of police sizes in neighboring states (*NEIGHBOR POLICE*) as instrumental variables for that state' police size. The fixed-effects GMM estimates show that police size reduces crime significantly in the U.S. More importantly, despite differences between our country sample and our U.S. sample (e.g., the frequency of observations and the definitions of household income), the two samples produce virtually identical results regarding our two main hypotheses.

Although not reporting in separate tables for brevity, we conduct various robustness checks about the current findings. As for the results in Table 2, we further disaggregate violent crime into several subcategories (homicide, robbery, assault, and rape) and find very similar results with robbery, assault and homicide, but not with rape. As for the results in Table 4, separate analyses for violent and property crimes still give very similar results. Finally, for both the country and the U.S. state level data, we conduct similar exercises with the dependent variable replaced by the labor force participation rate and find that the bipolarization index, not the conventional Gini index, reduces the unit's labor force participation rate, with other things being constant.

One consistent pattern we have observed across all the tables, but not explicitly mentioned, is that the explanatory power of the new bipolarization index is somewhat smaller at $\theta = 0$ than at $\theta = 0.25$. This is consistent with our previous conjecture: With the degree of income mobility being held constant, even high income earners should be considered in designing the bipolarization index as a crime determinant, although less so than low income earners. This motivates us to further explore the 'optimal' value of θ that gives our bipolarization index the greatest power in explaining crime rates. To that effect, we re-estimate each regression model in Tables 2, 3 and 4 with the entire set of control variables by fixed-effects GMM, repeatedly changing the value of θ from 0 to 0.5 by 0.01. For all cases, the estimated coefficients on the bipolarization index appears as a concave function of θ. When country data are used, the largest coefficient estimate of our bipolarization index is obtained at $\theta = 0.30$ and 0.32 for violent and total crime, respectively. For the U.S. data, the global optimum is obtained at $\theta = 0.27$. Due to the large standard error estimates, however, differences in the estimated coefficient for different values of θ are generally insignificant.

4 Conclusion

When a person decides whether to supply his labor in the legal or illegal labor market, he considers his relative position not just in the current income distribution but also in the distribution of expected lifetime income. While the former is related with conventional static inequality measures such as the Gini coefficient, the latter is determined by the person's entire lifetime income stream and his expected income mobility. Given this perspective, even a low income earner at a certain point in time would not have high crime incentive if he had better prospects in the future: if one has a higher expectation of upward mobility, then expected lifetime income is higher and so is the marginal cost of the crime.

The level of mobility implied by income distribution can be well summarized by income polarization. The index by Esteban and Ray (1994) or Esteban, Gradin and Ray (2007), for example, has strong implications on income mobility between groups, although such an aspect is not explicitly emphasized in their work. They specify that income distribution is more polarized when either the between-group distance is larger or within-group clustering is stronger. Increasing bipolarization reflects the phenomena of the disappearing middle class and the enhanced immobility between groups. However, such an immobility measure only describes the distributional aspect of the income profile of a society, which naturally assumes that all income groups are symmetric in the degree of alienation one group feels against the others. In order to find a more meaningful relation between income immobility and the incentive for crime, we need to introduce group heterogeneity in the feeling of alienation so that we can better capture overall level of social tension.

By developing a new polarization measure that accounts for between-group mobility and asymmetries of antagonism, we show that the bipolarization of income distribution is a crucial aspect in understanding crime incentive. The fixed effects GMM estimates for the cross country and the U.S. panel data regression models suggest that, while the Gini index (i.e., the income inequality measure) is not related with any type of crime, bipolarization significantly increases the crime rate. More importantly, consistent with our prediction regarding asymmetric feelings of alienation 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.

The asymptotic distribution of the proposed bipolarization index is also derived using results from U-statistics and an easy-to-implement jackknife variance estimation algorithm is obtained. As a result, more statistical inferences are expected in the future on cross-country comparisons or time changes of income polarization.

Appendix: Jackknife Variance Estimation

This appendix summarizes jackknife estimation of $V_{\mathcal{B}}(\alpha,\theta)$ in Theorem 2. It is known that bootstrap variance estimation of the Gini coefficient is still computationally demanding especially when n is large like the conventional income data. This is still the case for $\widehat{\mathcal{B}}_n(\alpha,\theta)$ since we need to calculate the Gini coefficients $\widehat{G}$ and $\widehat{G}_k$ 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 the following efficient algorithm. 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 in two using a given cutoff point y^* (e.g., the sample mean $\overline{y}$), and let $A_1 = \{y_i | y_i < y^*\}$ and $A_2 = \{y_i | y_i \geq y^*\}$.

   (a) Since the original data is already sorted in step 1, the data in 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 $\widehat{\pi}_k = n_k/n$ 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, calculate $\widehat{\mathcal{B}}_n(\alpha,\theta)$ as in (5) for given α and θ.

3. From the entire sample, 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

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

(b) Let

$$\widehat{\pi}_{1,(-i)} = \begin{cases} (n_1 - 1) / (n-1) & \text{if } y_i \in A_1 \\ n_1 / (n-1) & \text{if } y_i \in A_2 \end{cases}$$

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

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

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

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

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

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

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

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

$$V_i = V_{i-1} + \frac{n-1}{n} \left(\widehat{\mathcal{B}}_{n,(-i)}(\alpha, \theta) - \widehat{\mathcal{B}}_n(\alpha, \theta)\right)^2$$

with $V_0 = 0$. Then, V_n is the jackknife variance estimate of $\widehat{\mathcal{B}}_n(\alpha, \theta)$.

References

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

Alesina, A., R. Di Tella and R. MacCulloch (2004). Inequality and happiness: Are Europeans and Americans different?, *Journal of Public Economics*, 88, 2009-2042.

Atkinson, A.B. (1970). On the Measurement of Inequality, *Journal of Economic Theory*, 2, 244-263.

Atkinson, A.B., L. Rainwater, and M. Smeeding (1995). *Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study*, OECD Social Policy Studies No. 18, Paris: OECD.

Becker, G.S. (1963). Crime and Punishment: An Economic Approach, *Journal of Political Economy*, 76, 169-217.

Bishop, J.A., J.P. Formby, and B. Zheng (1997). Statistical inference and the Sen index of poverty, *International Economic Review*, 38, 381-388.

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.

Cowell, F.A. and K. Kuga (1981). Inequality Measurement. An Axiomatic Approach, *European Economic Review*, 15, 287-305.

Dagum, C. (1997). A New Approach to the Decomposition of the Gini Income Inequality Ratio, *Empirical Economics*, 22, 515-531.

Demombynes, G., and B. Özler (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.-Y., 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, 521-565.

Esteban, J., C. Gradin and D. Ray (2007). An Extension of a Measure of Polarization, with an Application to the Income Distribution of Five OECD Countries, *Journal of Economic Inequality*, 5, 1-19.

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

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

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.

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.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions, *Annals of Statistics*, 19, 293–325.

Imrohoroglu A., A. Merlo, and P. Rupert (2000). On the Political Economy of Income Redistribution and Crime, *International Economic Review*, 41, 1-25.

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 and Crime, *Review of Economics and Statistics*, 82, 530-539.

Lasso de la Vega, C., and A.M. Urrutia (2006). An Alternative formulation of Esteban-Gradin-Ray extended measure of polarization, *Journal of Income Distribution*, 15, 42-54.

Merton, R.K. (1968). *Social Theory and Social Structure*, New York: Free Press.

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

Shorrocks, A.F. (1984). Inequality decomposition by population subgroups, *Econometrica*, 52, 1369-1385.

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

Table 1. Determinants of Violent Crime (cross country data)

Variables	Basic Model $\theta = 0.5$	Basic Model $\theta = 0.25$	Basic Model $\theta = 0$	Extended Model $\theta = 0.5$	Extended Model $\theta = 0.25$	Extended Model $\theta = 0$
BIPOLARIZATION	1.654	11.969**	7.364**	0.646	7.469*	4.993**
	(11.150)	(5.026)	(2.731)	(3.986)	(3.519)	(2.012)
GINI	1.638	-2.587	0.102	0.342	-2.073	-0.286
	(4.916)	(2.374)	(1.790)	(2.094)	(1.440)	(1.078)
DENSITY	–	–	–	0.023	0.030	0.033*
				(0.022)	(0.019)	(0.018)
URBAN POPULATION	–	–	–	0.150***	0.122**	0.113**
				(0.037)	(0.042)	(0.041)
YOUNG MEN	–	–	–	2.038	2.146*	2.233*
				(1.178)	(1.073)	(1.095)
UNEMPLOYMENT	–	–	–	0.025	0.017	0.020
				(0.028)	(0.028)	(0.026)
EDUCATION	–	–	–	-0.055	-0.080	-0.100
				(0.066)	(0.067)	(0.064)
CONSTANT	-8.006	-6.808***	-6.418***	-8.006	-6.808***	-6.418***
	(1.452)	(1.472)	(1.452)	(1.452)	(1.472)	(1.452)
No. of Countries	13					
Obs.	68					

NOTE: Fixed-effects estimation. Dependent variable = log(number of violent crimes per 100,000 people), where violent crime includes robbery, homicide, assault and rape. θ is the weight parameter in the bipolarization index. Robust standard error estimates are in parentheses. *, ** and *** means significant at the 10%, 5%, and 1% levels, respectively.

Table 2. Determinants of Violent Crime - Consideration of Crime Prevention Activity (cross country data)

	Weight on High Income Group		
	$\theta = 0.5$	$\theta = 0.25$	$\theta = 0$
Exogenous police size (fixed effects estimates)			
BIPOLARIZATION	1.891	8.195**	5.529**
	(4.982)	(3.740)	(2.122)
GINI	-0.743	-2.767*	-0.750
	(2.878)	(1.548)	(0.949)
log(POLICE)	0.374*	0.319	0.287
	(0.173)	(0.206)	(0.213)
Endogenous police size (fixed effects GMM estimates)			
Instrumental variable: TAX-GDP-RATIO			
BIPOLARIZATION	-4.690	11.258	7.598**
	(12.258)	(7.136)	(3.544)
GINI	2.104	-4.757	-1.808
	(7.906)	(4.033)	(2.161)
log(POLICE)	-3.235	-3.049	-2.373
	(4.209)	(4.152)	(3.194)
First-stage F-test	0.165	0.042	0.019
No. of Countries		13	
Obs.		60	

NOTE: Dependent variable = log(number of violent crimes per 100,000 people), where violent crime includes robbery, homicide, assault and rape. Additional control variables are as in Table 1. θ is the weight parameter in the bipolarization index. Robust standard error estimates are in parentheses. *, ** and *** means significant at the 10%, 5%, and 1% levels, respectively. 'First-stage F-test' gives the p-value.

Table 3. Determinants of Total Crime - Consideration of Crime Prevention Activity (cross country data)

	Weight on High Income Group		
	$\theta = 0.5$	$\theta = 0.25$	$\theta = 0$
Exogenous police size (fixed effects estimates)			
BIPOLARIZATION	0.270	2.953*	1.960**
	(2.928)	(1.376)	(0.742)
GINI	0.305	-0.694	-0.694
	(1.846)	(0.979)	(0.816)
log(POLICE)	0.357	0.351	0.339
	(0.251)	(0.261)	(0.267)
Endogenous police size (fixed effects GMM estimates)			
Instrumental variables: TAX-GDP-RATIO			
CRIME COMPOSITION			
BIPOLARIZATION	-2.817	2.381*	1.889**
	(3.005)	(1.332)	(0.757)
GINI	2.272	-0.192	0.312
	(0.312)	(0.814)	(0.662)
log(POLICE)	0.374	0.205	0.100
	(0.356)	(0.393)	(0.400)
First-stage F-test	0.000	0.000	0.000
Hansen J-test	0.021	0.030	0.055
No. of Countries		13	
Obs.		60	

NOTE: Dependent variable = log(total crime per 100,000 people). Additional control variables are as in Table 1. θ is the weight parameter in the bipolarization index. Robust standard error estimates are in parentheses. *, ** and *** means significant at the 10%, 5%, and 1% levels, respectively. ‘First-stage F-test’ and ‘Hansen J-test’ give the p-values.

Table 4. Determinants of Total Crime (U.S. state-level data)

	Weight on High Income Group		
	$\theta = 0.5$	$\theta = 0.25$	$\theta = 0$
Exogenous police size (fixed effects estimates)			
BIPOLARIZATION	-0.242	0.519*	0.415**
	(0.513)	(0.307)	(0.191)
GINI	-0.816**	-1.459***	-1.530***
	(0.306)	(0.302)	(0.275)
log(POLICE)	-0.095	-0.097	-0.099
	(0.072)	(0.071)	(0.071)
Endogenous police size (fixed effects GMM estimates)			
Instrumental variables: log(NEIGHBOR POLICE)			
CRIME COMPOSITION			
BIPOLARIZATION	-0.484	0.595**	0.494***
	(0.499)	(0.259)	(0.166)
GINI	-0.436	-1.286***	-1.480***
	(0.318)	(0.281)	(0.286)
log(POLICE)	-0.946***	-0.940***	-1.392***
	(0.231)	(0.231)	(0.263)
First-stage F-test	0.000	0.000	0.000
Hansen J-test	0.234	0.253	0.282
No. of State		40	
Obs.		600	

NOTE: Dependent variable = log(total crime per 100,000 people). Additional control variables are the ratio of (the number of young men of age 15-29)/(the total population), state unemployment rates, the population proportion of college graduates, and population density. θ is the weight parameter in the bipolarization index. Robust standard error estimates are in parentheses. *, ** and *** means significant at the 10%, 5%, and 1% levels, respectively. 'First-stage F-test' and 'Hansen J-test' give the p-values.

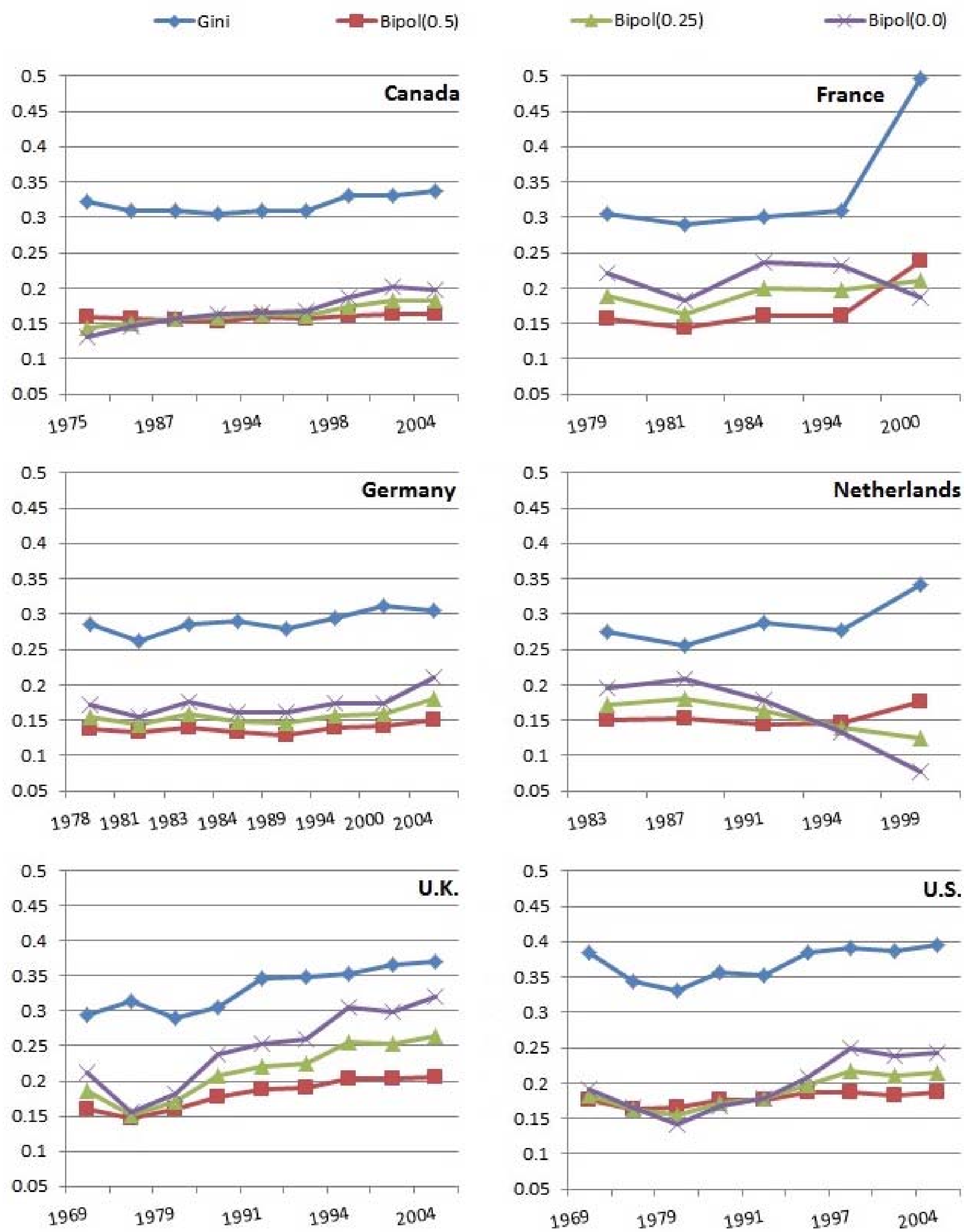

Figure 2: Inequality and Bipolarization (θ values are in parentheses; $\alpha = 1.6$ for all cases)