Distributional Overlap: Simple, Multivariate, Parametric and Nonparametric Tests for Alienation, Convergence and General Distributional Difference Issues

Gordon Anderson*
University of Toronto

Ying Ge †
University of International Business and Economics

Teng Wah Leo‡
St. Francis Xavier University

February 27, 2009

*Department of Economics, University of Toronto. Email Address: anderson@chass.utoronto.ca. Part of this work was carried out in part whilst the first author was visiting the Institute For Fiscal Studies and has benefited from the help and advice of Alissa Goodman, Ian Crawford and participants at the Macro Seminar at Wisconsin Madison and the Economics Seminar at Southern Methodist University . Support from the Institute and the SSHRC under grant number 410040254 is gratefully acknowledged.

†School of International Trade and Economics, University of International Business and Economics.

‡Department of Economics, St. Francis Xavier University. Email Address: tleo@stfx.ca

1 Introduction

In keeping with the Akerlof (1997) notion of social distance the terminologies of *Polarization*, *Alienation* and *Convergence* are gaining an ever increasing currency in Economics (Anderson (2005) provides a limited list of the usage). The terms have to do with the extent to which agents in identifiable groups *Polarize* from each other, by finding increasing within group affinity (*Convergence*) and decreasing between group affinity (*Alienation*). As Atkinson (1998) stresses the phenomena are inherently multi-dimensional so that formal measures will depend upon aggregated distances in multivariate space between the economic variables of that agent and those of the rest of society. The trick is to develop simple expedient tests that capture this type of phenomenon.

Esteban and Ray (1994) (hereafter referred to as ER) and Duclos, Esteban, and Ray (2004) (hereafter referred to as DER) provided indices which identify the phenomenon in the mixture of group distributions in a univariate framework. Anderson (2004a) and Anderson (2004b), in studying the anatomy of polarized states, provided tests (again in a univariate framework) which explore the anatomical features of polarizing phenomena both in terms of observable group distributions and in terms of their implications for mixtures of those distributions. It transpires that polarization is not just simply a case of reduced within group variances and increased between group distance, it can occur with constant within group variances and between group locations when the groups exhibit mean and variance preserving appropriate skewing patterns. However whenever polarization presents itself, except in one pathological case, it is invariably associated with diminished (or at least not increased) distributional overlap.

The degree of overlap measures the points of "commonality", "likeness" or "coherence" between two groups, the extent to which they do not overlap records an index of the alienation of the group. It is unlike deprivation in the sense that groups with a surfeit of goods relative to the rest of society can be considered alienated just as those with a deficit can. It is unlike the ER and DER definition in that two identical within group economic variable distributions will yield 0 alienation under the overlap measure but will yield increasing alienation under the ER and DER measure, because of the within group increased dissociation[1].

[1] Alienation tests have long existed in the econometrics literature, see for example Dhrymes (1970). Typically, given co-varying vectors $\mathbf{y}$ and $\mathbf{x}$ of dimension $m \times 1$ and $n \times 1$ respectively with a conformably

Weitzman (1970) first proposed a nonparametric version of the overlap measure[2] for the one dimensional case where p.d.f's intersected but once. More generally when two smooth, continuous distributions $f(x, y, z, ..)$ and $g(x, y, z, ..)$ have multiple intersections, the Overlap Measure or Index OI we introduce here may be written as:

$$OI = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \min\{f(x, y, z, ...), g(x, y, z, ...)\} dxdydz... \tag{1}$$

It is clear that OI is bounded between 0 and 1 corresponding to no overlap and "complete" overlap respectively, consequently $IA = 1 - OI$ provides an index of the degree to which two populations are *polarized* or *alienated.* Here discussion will continue in terms of the univariate case though examples of multivariate situation will be presented.

Deutsch and Silber (1997) first proposed a similar Overlap Index to our nonparametric approach, and showed its relationship to the Pietra Index and Gini's Concentration Ratio. The main contribution of our work here is in showing the limiting distribution of the Overlap Index for both the parametric and our nonparametric Overlap Index, thereby permitting inferences. Stine and Heyse (2001) proposed that the unknown density in the Overlap Index could be estimated by kernel densities. We have refrained from using this method as our focus is to provide a Overlap Index that is amenable to inferences and hypothesis testing, since the limiting distribution of the Overlap Index using the kernel density approach is unknown, but is the subject of future work.

In the following the parametric Overlap Index and its distribution is outlined in Section 2, Section 3 discusses issues concerning the implementation of the nonparametric version, three examples illustrating uses of the index are proffered in section 4 and section 5 concludes.

---

partitioned covariance matrix Σ of the form:

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yx}} \\ \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xx}} \end{bmatrix}$$

such tests focus on the ratio $\frac{|\Sigma|}{(|\Sigma_{\mathbf{yy}}||\Sigma_{\mathbf{xx}}|)}$ and highlight the lack of covariation between $\mathbf{y}$ and $\mathbf{x}$ rather than the lack of similarity between the marginal distributions of $\mathbf{y}$ and $\mathbf{x}$.

[2]The Weitzman (1970) overlap measure is of the form,

$$\left[\int_{\min(x|f(x))}^{x^*} f(x)dx + \int_{x^*}^{\max(x|g(x))} g(x)dx \right]$$

where x^* is the unique point at which $f(x)$ and $g(x)$ intersect.

2 The Parametric Overlap Index

Let $\mathbf{x}^f \in \mathbf{X}^f$ and $\mathbf{x}^g \in \mathbf{X}^g$, be two random variables, where $\mathbf{X}^f$ and $\mathbf{X}^g \subset \mathbb{R}$ denote the support of the respective random variables. Let their respective continuous probability density functions be $f(\mathbf{x}^f, \theta^f)$ and $g(\mathbf{x}^g, \theta^g)$, where θ^f and θ^g are the respective $J \times 1$ parameter vectors of the density functions $f(.,.)$ and $g(.,.)$. Then by the above assumptions, $\frac{\partial \pi_k^i(\theta^i)}{\partial \theta_j^i}$, where $k \in \{1, 2, ..., K+1\}$ and $j \in \{1, 2, ..., J\}$, exists. Partitioning the range of $\mathbf{x}^i$, where $i \in \{f, g\}$, into $K+1$ *mutually exclusive and exhaustive intervals*, defined by K *common* partition points (*common* for both $f(\mathbf{x}^f, \theta^f)$ and $g(\mathbf{x}^g, \theta^g)$), let π^i be the $(K+1) \times 1$ vector of true probabilities, with typical elements denoted by $\pi_k^i(\theta^i) \equiv \pi_k^i$, where $i \in \{f, g\}$ and $k \in \{1, 2, ..., K+1\}$, which is the true probability of a realization of $\mathbf{x}^i$ being in partition k. Then given a sample of size n^i, $i \in \{f, g\}$, drawn from the respective populations, with each observation denoted as $\mathbf{x}_l^i$, $l \in \{1, 2, ..., n^i\}$, and consistent and asymptotically efficient estimates of the $J \times 1$ parameter vectors θ^f and θ^g, denote $\widehat{\pi}_k^i$ as the estimator of π_k^i, where $i \in \{f, g\}$ and $k \in \{1, 2, ..., K\}$.

Let $\mathbf{d}^i$ be a $(K+1) \times 1$ vector whose typical k'th element, $d_k^i = \sqrt{n^i}\left(\frac{\widehat{\pi}_k^i - \pi_k^i}{\sqrt{\pi_k^i}}\right)$, where $i \in \{f, g\}$. Next, let $\mathbf{M}_i$ be a $(K+1) \times J$ matrix of rank J, with the (k, j)'th element being $\frac{1}{\sqrt{\pi_k^i}} \frac{\partial \pi_k^i}{\partial \theta_j^i}$, where $i \in \{f, g\}$. Finally, let $\mathbf{M}_i'\mathbf{M}_i = \mathcal{H}_i$, and

$$\Omega_i = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & 1 \end{bmatrix} - \begin{bmatrix} \sqrt{\pi_1^i} \\ \sqrt{\pi_2^i} \\ \vdots \\ \sqrt{\pi_{K+1}^i} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_1^i} & \sqrt{\pi_2^i} & ... & \sqrt{\pi_{K+1}^i} \end{bmatrix}$$

Then from Rao (1973) pages 383 and 392, the asymptotic distribution of $\mathbf{d}^i$ may be written as,

$$\mathbf{d}^i \overset{a}{\sim} N_{K+1}(\mathbf{0}, (\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i')'\Omega_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i')) \tag{2}$$

Let

$$\mathbf{B} = \begin{bmatrix} \sqrt{\pi_1^i} & 0 & ... & 0 \\ 0 & \sqrt{\pi_2^i} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \sqrt{\pi_{K+1}^i} \end{bmatrix}$$

Then equation (2) can be written as,

$$\sqrt{n^i}\left(\widehat{\pi}^i - \pi^i\right) \overset{a}{\sim} N_{K+1}(\mathbf{0}, \mathbf{B}_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i)'\Omega_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i)\mathbf{B}_i') \tag{3}$$

$$\Rightarrow \widehat{\pi}^i \overset{a}{\sim} N_{K+1}\left(\pi^i, (\mathbf{B}_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i)'\Omega_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i)\mathbf{B}_i')\frac{1}{n^i}\right) \tag{4}$$

where both π^i and $\widehat{\pi}^i$, $i \in \{f, g\}$, are $(K+1) \times 1$ vectors. Let $\mathbf{i}$ be a $(K+1) \times 1$ vector of ones. Then it follows that,

$$\Rightarrow \mathbf{i}'\widehat{\pi}^i \overset{a}{\sim} N\left(\mathbf{i}'\pi, (\mathbf{i}'\mathbf{B}_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i)'\Omega_i(\mathbf{M}_i\mathcal{H}_i^{-1}\mathbf{M}_i)\mathbf{B}_i'\mathbf{i})\frac{1}{n^i}\right) \tag{5}$$

Let $\widehat{\pi}^{\min} = \min\{\widehat{\pi}^f, \widehat{\pi}^g\}$, and define the estimate of OI as $\widehat{OI} = \mathbf{i}'\widehat{\pi}^{\min}$. Then from equations (4) and (5),

$$\widehat{OI} \overset{a}{\sim} N\left(\mathbf{i}'\pi^{\min}, (\mathbf{i}'\mathbf{B}_{\min}(\mathbf{M}_{\min}\mathcal{H}_{\min}^{-1}\mathbf{M}_{\min})'\Omega_{\min}(\mathbf{M}_{\min}\mathcal{H}_{\min}^{-1}\mathbf{M}_{\min})\mathbf{B}_{\min}'\mathbf{i})\frac{1}{n^{\min}}\right) \tag{6}$$

It is clear from equation (6) that as $K \to \infty$ (but $\frac{K}{n^{\min}} \to 0$), $\widehat{OI} \to OI$. Alternatively if the K partition points are defined by the intersection points of $f(.)$ and $g(.)$ which, given consistent estimates of θ^f and θ^g respectively, can be consistently estimated by the solution or solutions to $f(\mathbf{x}^*, \theta_f) = g(\mathbf{x}^*, \theta_g)$, then $\widehat{OI} \to OI$ as $n^{\min} \to \infty$.

For independently sampled observations in states 1 and 2, changes in the degree of convergence can be examined by focusing on $\Delta\widehat{OI} = \widehat{OI}_1 - \widehat{OI}_2$ which, under the null of no change and with the respective variances defined by equation (6) and written as V_1 and V_2, will be distributed as:

$$\Delta\widehat{OI} \overset{a}{\sim} N(0, V_1 + V_2) \tag{7}$$

At the expense of some complexity, comparisons of non-independent distributions can be made in a similar fashion by including the covariances between the two distributions.

3 The Nonparametric Index

For the single dimension case, for K known or predetermined partition points y_k, $k \in \{1, .., K\}$ (where for convenience assume that the partition points are in ascending rank order, $y_1 < y_2 < ... < y_K$)[3] the Overlap Index may be implemented empirically by:

$$\widehat{OI} = \left\{ \begin{array}{cc} & \max\left(\sum_{r=1}^{n^f} \frac{\mathbb{I}(\mathbf{x}_r^f - y_k)}{n^f}, \sum_{s=1}^{n^g} \frac{\mathbb{I}(\mathbf{x}_s^g - y_k)}{n^g} \right) \\ - & \sum_{k=1}^{K} \min\left(\sum_{r=1}^{n^f} \frac{\mathbb{I}(\mathbf{x}_r^f - y_k) - \mathbb{I}(\mathbf{x}_r^f - y_{k-1})}{n^f}, \sum_{s=1}^{n^g} \frac{\mathbb{I}(\mathbf{x}_s^g - y_k) - \mathbb{I}(\mathbf{x}_s^g - y_{k-1})}{n^g} \right) \end{array} \right\} \tag{8}$$

where $\mathbf{x}_r^f$, $r \in \{1, 2, ..., n^f\}$, corresponds to the realization of observation r of random variable $\mathbf{x}^f$ associated with the density function $f(.)$ and $\mathbf{x}_s^g$, $s \in \{1, 2, ..., n^g\}$, corresponds to the realization of observation s of random variable $\mathbf{x}^g$ associated with the density function $g(.)$. $\mathbb{I}(.)$ is an indicator function where $\mathbb{I}(z) = 1$ if $z \leq 0$ and is 0 otherwise. In addition, y_0 is chosen such that $\mathbb{I}(\mathbf{x}_r^f - y_0) = \mathbb{I}(\mathbf{x}_s^g - y_0) = 0$, for all $r \in \{1, 2, ..., n^f\}$ and $s \in \{1, 2, ..., n^g\}$. When the samples on $\mathbf{x}^f$ and $\mathbf{x}^g$ are independent, the asymptotic distribution of the index is well defined (See Rao (1973) page 383, result (i)). To see this, first denote as before π^i and $\widehat{\pi}^i$, $i \in \{f, g\}$, as the $(K+1) \times 1$ vector of true and estimated probabilities respectively. The typical element of $\widehat{\pi}^i$, $i \in \{f, g\}$ is given by the proportion of sample elements falling in the $K+1$ intervals defined by the partition points y_k, $k \in \{1, 2, ..., K\}$. Then the asymptotic distribution of $\widehat{\pi}^f$ is:

$$\widehat{\pi}^f \overset{a}{\sim} N_{K+1}\left(\pi^f, \left(\mathrm{dg}(\pi^f) - \pi^f \pi^{f\prime}\right) \frac{1}{n^f}\right) \tag{9}$$

Similarly for $\widehat{\pi}^g$:

$$\widehat{\pi}^g \overset{a}{\sim} N_{K+1}\left(\pi^g, \left(\mathrm{dg}(\pi^g) - \pi^g \pi^{g\prime}\right) \frac{1}{n^g}\right) \tag{10}$$

Note that when $g(.)$ is the stochastically dominated distribution, the vector $\widehat{\pi}^f - \widehat{\pi}^g$ will alternate in sign with the first element being negative. As before, let $\widehat{\pi}^{\min} = \min\{\widehat{\pi}^f, \widehat{\pi}^g\}$, and define the estimate of OI as $\widehat{OI} = \mathbf{i}'\widehat{\pi}^{\min}$. Then the asymptotic distribution of $\widehat{\pi}^{\min}$ is,

$$\widehat{\pi}^{\min} \overset{a}{\sim} N_{K+1}\left(\pi^{\min}, \left(\mathrm{dg}(\pi^{\min}) - \pi^{\min} \pi^{\min\prime}\right) \frac{1}{n^{\min}}\right) \tag{11}$$

where $n^{\min}$ alternates with the appropriate sample size deflator n^f or n^g. The asymptotic distribution of $\widehat{OI}$ is,

$$\widehat{OI} := \widehat{\pi}^{\min} \overset{a}{\sim} N\left(\mathbf{i}'\pi^{\min}, \left(\mathbf{i}'\left(\mathrm{dg}(\pi^{\min}) - \pi^{\min} \pi^{\min\prime}\right)\mathbf{i}\right) \frac{1}{n^{\min}}\right) \tag{12}$$

[3]Like the classic Goodness of Fit test, these instruments when employed in comparison tests encounter the criticism that the tests are inconsistent (Barrett and Donald 2003). The issue really only arises when too few partitions are chosen as Appendix 1 demonstrates.

On the other hand, if the researcher is interested in examining the degree of alienation or polarization between two populations (Anderson 2003, Anderson 2004a, Esteban and Ray 1994, Duclos, Esteban, and Ray 2004), the appropriate measure should be $IA = 1 - OI$. Denote $\widehat{IA}$ as the estimate of IA, then

$$\widehat{IA} = 1 - \widehat{OI} = 1 - \mathbf{i}'\widehat{\pi}^{\min} \tag{13}$$

It is clear that the asymptotic distribution of $\widehat{IA}$ is,

$$\widehat{IA} \overset{a}{\sim} N\left(1 - \mathbf{i}'\pi^{\min}, \mathbf{i}'\left(\left(\mathrm{dg}(\pi^{\min}) - \pi^{\min}\pi^{\min\prime}\right)\frac{1}{n^{\min}}\right)\mathbf{i}\right) \tag{14}$$

In the literature on *Polarization*, it is often of interest to examine the evolution of polarization across periods or borders. The above result consequently allows for inference. Consider now two independently sampled observations in states 1 and 2, then the change in the degree of alienation can be examined by focusing on the difference in IA, $\Delta IA = IA_1 - IA_2$, where IA_m, $m \in \{1, 2\}$, denote the alienation measure IA in state m. Denoting the estimate of IA_m as $\widehat{IA}_m$, and the estimate of ΔIA as $\Delta\widehat{IA}$, then under the null of no change in alienation, $\Delta\widehat{IA}$ will be distributed as:

$$\Delta\widehat{IA} \overset{a}{\sim} N\left(0, \mathbf{i}'\left(\left(\mathrm{dg}(\pi^{\min,1}) - \pi^{\min,1}\pi^{\min,1\prime}\right)\frac{1}{n^{\min,1}}\right)\mathbf{i} + \mathbf{i}'\left(\left(\mathrm{dg}(\pi^{\min,2}) - \pi^{\min,2}\pi^{\min,2\prime}\right)\frac{1}{n^{\min,2}}\right)\mathbf{i}\right) \tag{15}$$

where $n^{\min,1}$ and $n^{\min,2}$ are the relevant sample size deflators from states 1 and 2 respectively, as defined in (11). The asymptotic distribution of $\widehat{IA}$ in the parametric case can be similarly derived.

In practice, the nonparametric multi-dimensional case will soon run into the curse of dimensionality that bedevils nonparametric estimation, however its implementation simply demands that the vectors π^f, π^g and $\pi^{\min}$, and $\widehat{\pi}^f$, $\widehat{\pi}^g$ and $\widehat{\pi}^{\min}$ correspond to the vectorized list of cell proportions and probabilities generated by partitions of the support of the respective underlying random variables. Of course if intersections are to be used in determining cells, it is no longer a question of estimating intersection points, but one of estimating intersection functions.

3.1 Practical Considerations in Implementing the Non-Parametric Test

Often theory will prescribe the parametric nature of distributions being considered (such as in example 1) in which case parametric overlap tests provide a useful tool for evaluating theory. Frequently nonparametric versions of the overlap test will have to be relied upon in the absence of theoretical guidance. Parametric versions of the index estimates of the intersection points and the index are at least asymptotically unbiased, but the nonparametric versions we propose in this paper are not.

There are two sources of bias in the estimate of the nonparametric overlap measure which work in opposing directions. The first source is the bias induced by estimating the intersection points[4] and is associated with intervals which span the true intersection points, exaggerating the degree of overlap (hence understating the degree of alienation).

This is best illustrated by considering a simple example with two random variables, $\mathbf{x}^f$ and $\mathbf{x}^g$, defined on the support of $\mathbf{X}^f$ and $\mathbf{X}^g$, where both $\mathbf{X}^f$ and $\mathbf{X}^g \subset \mathbb{R}$. Without loss of generality, suppose $\mathbf{X}^f = [\underline{\mathbf{x}}, \overline{\mathbf{x}}] = \mathbf{X}^g$, in other words the two random variables have common support. Further suppose the density functions intersect only once at $y^* \in (\underline{\mathbf{x}}, \overline{\mathbf{x}})$ such that $\min\{f(\mathbf{x}^f), g(\mathbf{x}^g)\} = g(\mathbf{x}^g)$ for any $\mathbf{x}^f$, $\mathbf{x}^g \in [\underline{\mathbf{x}}, y^*)$, and $\min\{f(\mathbf{x}^f), g(\mathbf{x}^g)\} = f(\mathbf{x}^f)$ for any $\mathbf{x}^f$, $\mathbf{x}^g \in [y^*, \overline{\mathbf{x}}]$. Then the true overlap measure is,

$$OI = \int_{\underline{\mathbf{x}}}^{y^*} g(\mathbf{x}^g) d\mathbf{x}^g + \int_{y^*}^{\overline{\mathbf{x}}} f(\mathbf{x}^f) d\mathbf{x}^f \tag{16}$$

Consider now the estimate of y^* denoted as $\widehat{y}^* \neq y^*$. Without loss of generality, suppose

[4]Generally the intersection points will not be known, however they could be estimated by considering kernel estimates of $f(x) - g(x)$ for δ size incremental values of x and, when the sign changed between x and $x + \delta x^*$, the point at which the functions intersect can be estimated by:

$$x^* = \frac{(|f(x+\delta) - g(x+\delta)| \times x + |f(x) - g(x)| \times (x+\delta))}{(|f(x) - g(x)| + |f(x+\delta) - g(x+\delta)|)}$$

$\widehat{y}^* > y^*$. Then the estimate of the Overlap Index is,

$$\begin{aligned}
\widehat{OI} &= \int_{\underline{\mathbf{x}}}^{\widehat{y}^*} g(\mathbf{x}^g)d\mathbf{x}^g + \int_{\widehat{y}^*}^{\overline{\mathbf{x}}} f(\mathbf{x}^f)d\mathbf{x}^f \\
&= \int_{\underline{\mathbf{x}}}^{y^*} g(\mathbf{x}^g)d\mathbf{x}^g + \int_{y^*}^{\widehat{y}^*} g(\mathbf{x}^g)d\mathbf{x}^g + \int_{\widehat{y}^*}^{\overline{\mathbf{x}}} f(\mathbf{x}^f)d\mathbf{x}^f \\
&= \int_{\underline{\mathbf{x}}}^{y^*} g(\mathbf{x}^g)d\mathbf{x}^g + \int_{y^*}^{\overline{\mathbf{x}}} f(\mathbf{x}^f)d\mathbf{x}^f + \left(\int_{y^*}^{\widehat{y}^*} g(\mathbf{x}^g)d\mathbf{x}^g - \int_{y^*}^{\widehat{y}^*} f(\mathbf{x}^f)d\mathbf{x}^f \right)
\end{aligned}$$

which then implies that,

$$\widehat{OI} = OI + \left(\int_{y^*}^{\widehat{y}^*} \left(g(\mathbf{x}^g) - f(\mathbf{x}^f)\right) d\mathbf{x}^g d\mathbf{x}^f \right) = OI + \text{Bias}^+ \tag{17}$$

So that in so far as $g(.)$ dominate $f(.)$, the bias is positive. It is easy to show that there is a positive bias when $y^* > \widehat{y}^*$ as well. In other words, the overlap measure, $\widehat{OI}$ overstates the true degree of overlap, which in turn imply that the alienation measure, $\widehat{IA}$ is understated.

The second source of bias is related to the $\min\{\widehat{\pi}^f, \widehat{\pi}^g\}$ function, where $\widehat{\pi}^f$ and $\widehat{\pi}^g$ are the respective vectors of independent estimates of the probability of being in an interval under $f(.)$ and $g(.)$, with elements denoted $\widehat{\pi}^i_k$, for $i \in \{f, g\}$ and $k \in \{1, 2, ..., K+1\}$, and understates the degree of overlap (hence exaggerating the degree of alienation). This bias derives from the fact that in general for independent random variables $\widehat{\pi}^f_k$ and $\widehat{\pi}^g_k$, $E(\widehat{\pi}^f_k) \geq E(\widehat{\pi}^f_k | \widehat{\pi}^f_k < \widehat{\pi}^g_k)$. Since $\widehat{\pi}^f_k$ is an unbiased estimator for $f(.)$, the conditional estimator implicit in the $\min\{.,.\}$ function will be downward biased. Generally for independent $\widehat{\pi}^f_k$ and $\widehat{\pi}^g_k$, with respective p.d.f's $p(\widehat{\pi}^f_k)$ and $q(\widehat{\pi}^g_k)$, and respective c.d.f's $P(\widehat{\pi}^f_k)$ and $Q(\widehat{\pi}^g_k)$, $E(\widehat{\pi}^f_k | \widehat{\pi}^f_k < \widehat{\pi}^g_k)$ is given by:

$$\begin{aligned}
\int_0^1 \int_{\widehat{\pi}^f_k}^1 \widehat{\pi}^f_k p(\widehat{\pi}^f_k) q(\widehat{\pi}^g_k) d\widehat{\pi}^g_k d\widehat{\pi}^f_k &= \int_0^1 \widehat{\pi}^f_k p(\widehat{\pi}^f_k) \left(1 - Q(\widehat{\pi}^f_k)\right) d\widehat{\pi}^f_k \\
&= E(\widehat{\pi}^f_k) - \int_0^1 \widehat{\pi}^f_k p(\widehat{\pi}^f_k) Q(\widehat{\pi}^f_k) d\widehat{\pi}^f_k
\end{aligned} \tag{18}$$

Since the last term is never negative, the overlap measure will always be biased downward by this component and the alienation measure exaggerated as a consequence. Appendix 2 reports a small Monte Carlo study examining these separate effects.

4 Some Examples

This section illustrates the use of both the parametric and nonparametric approach to calculating the Overlap measure with three examples. The first example provides a univariate parametric application to the issue of divergent or convergent economic growth in Chinese cities with differing administrative structures. This example highlights how prior economic or statistical theoretical work can provide guidance as to the appropriate distribution to use in calculating the Overlap Index. In the absence of this guidance, the researcher would have to rely on the nonparametric Overlap Index of the latter examples. The second example illustrates the ease with which the measure can be obtained in the discrete univariate case through studying the impact of different family structures on the grade attainments of youths. The third example demonstrates its use in a continuous multivariate nonparametric environment, in considering the plight of single parent families and pensioners in the United Kingdom.

4.1 Example 1: Urban Income Size Distribution between Prefecture and County Level Cities in China

In 1978 China embarked upon a series of Economic Reforms which have had a profound impact on the Chinese economy, particularly on the urban economy and structure. The unique administrative hierarchical structure (which determines the relevant administrative and jurisdictional entity a city is subsumed within) of the Chinese Urban System was constructed in the command economy period of the 1950s to be compatible with the central planning system. Cities were largely of two administrative types; *Prefecture* level cities which had senior status and more power and *County* level cities which had junior level status. Prior to 1978 in the pre-reform period, most state manufacturing industries were located in or around the political centres which were generally *Prefecture* level cities, and the growth of investment in the state manufacturing sector was the main determinant of urban income growth. Consequently, the investment capacity of a city was closely related to its administrative level, with *Prefecture* level cities having much greater investment capacity and autonomy than *County* level cities. Inevitably a size and income disparity between these two types of cities was engendered and strict migration controls ensured this hierarchical structure remained stable in the pre-reform period.

The economic reforms presented challenges to this stability. Generally, the reforms involved political decentralization, economic liberalization, and openness to foreign trade and investment, and gradually changed the fundamental sources of urban growth. Firstly political decentralization delivered more economic power and autonomy to local governments, together with greater general policy autonomy helped reduce the "power" gap between the cities with different administrative status. Secondly economic liberalization, through stimulating rapid private sector growth and shrinking of the relative size of the state sector, favoured the *County* level cities which were characterized by relatively larger private sector and a greater dependence on foreign trade and investment. Thirdly, the large state sector share in *Prefecture* level cities, with its attendant inflexibilities, presented a distinct disadvantage in the transition process, while *County* level cities having a more malleable structure were more able to adjust quickly to fit into a market economy environment. The question then is whether these changes resulted in the urban income size distributions of these two city types becoming more alike. Following (Anderson and Ge 2009) data on per capita GDP for Chinese cities in 1990 and 1999 are used for the comparisons.

There are good theoretical reasons for believing that income size distributions are log normal (Gibrat 1930), however both Pareto and Double Pareto distributions also have a claim as candidates (See Pareto (1897), Champernowne (1953) and Reed (2001)). Table 1 reports goodness of fit tests and upper tail probabilities for the comparisons using all three distributions. The evidence for the log normal distribution is compelling since with the exception of the Prefecture/County mixture in 1990, the log normality specification is never rejected at the 1% level, whereas it is always rejected for all other specifications at sizes as low as 0.01%.

Let w^p and w^c denote income variables for prefecture and county level cities respectively, each assumed to be normally distributed, $w^p \sim N(\mu_p, \sigma_p)$ and $w^c \sim N(\mu_c, \sigma_c)$. Without loss of generality, let $\sigma_p > \sigma_c$. In addition, denote $\delta = \mu_p - \mu_c$ and $\sigma = \left(\frac{\sigma_p}{\sigma_c}\right)$. Firstly, since there is theoretical guidance relating the income distribution here, the partitions are defined by the intersection points of the two density functions. The intersection points for the two log normal densities are obtained from the solution to the equation,

$$\phi(\mu_p, \sigma_p) = \phi(\mu_c, \sigma_c) \tag{19}$$

where $\phi(.,.)$ denote the log normal density function. The solution is just,

$$\mu_c + \frac{\delta \pm \sigma\sqrt{\delta^2 - (1-\sigma^2)\ln\sigma^2}}{(1-\sigma^2)} \tag{20}$$

We can then denote the two intersection points as $y_k \equiv y_k(\mu_p, \mu_c, \sigma_p, \sigma_c)$, where $k \in \{1, 2\}$. Next, from our discussion in section 2, in order to obtain the standard error of the parametric

Table 1: Goodness of Fit Tests for Log-Normal, Pareto and Double Pareto Distributions

	1990		1999	
	Log Normal Distribution			
All cities	36.9483	(0.0000)	3.9231	(0.8640)
Prefectural	18.2626	(0.0193)	3.0678	(0.9300)
County	12.0877	(0.1473)	9.4637	(0.3047)
	Pareto Distribution			
All cities	4176.0	(0.0000)	5967.0	(0.0000)
Prefectural	1611.0	(0.0000)	2124.0	(0.0000)
County	2565.0	(0.0000)	3843.0	(0.0000)
	Double Pareto Distribution			
All Cities	1703.9	(0.0000)	2321.5	(0.0000)
Prefecural	597.9	(0.0000)	637.7	(0.0000)
County	968.8	(0.0000)	1491.9	(0.0000)

Note: Upper Tail probabilities in parenthesis

Prefecture and County level sample sizes in braces.

Overlap Index, we need to calculate the sample counterparts to the elements of the matrix $\mathbf{M}_i$ in equation (2), $i \in \{p, c\}$, with typical element being $\frac{1}{\sqrt{\pi_k^i}} \frac{\partial \pi_k^i}{\partial \theta_j}$, where $i \in \{p, c\}$, $k \in \{1, 2, 3\}$ and $j \in \{1, 2\}$, since in the case on hand, the parameters corresponding to the log normal density are just (μ_i, σ_i), $i \in \{p, c\}$. Next denoting $\Phi(.)$ as the log normal c.d.f., and noting that $\pi_k^i = \int_{y_{k-1}}^{y_k} \phi(w^i) dw^i$, the formulas for the partials for the respective parameters are,

$$\frac{\partial \pi_1^i}{\partial \mu_i} = \Phi\left(\frac{\partial y_1}{\partial \mu_i}\right) + \left(\mathbf{E}(w^i) - \mu_i\right) \frac{\pi_1^i}{\sigma_i} \tag{21}$$

$$\frac{\partial \pi_2^i}{\partial \mu_i} = \Phi\left(\frac{\partial y_2}{\partial \mu_i}\right) - \Phi\left(\frac{\partial y_1}{\partial \mu_i}\right) + \left(\mathbf{E}(w^i) - \mu_i\right) \frac{\pi_2^i}{\sigma_i} \tag{22}$$

$$\frac{\partial \pi_3^i}{\partial \mu_i} = \left(\mathbf{E}(w^i) - \mu_i\right) \frac{\pi_3^i}{\sigma_i} - \Phi\left(\frac{\partial y_2}{\partial \mu_i}\right) \tag{23}$$

and

$$\frac{\partial \pi_1^i}{\partial \sigma_i} = \Phi\left(\frac{\partial y_1}{\partial \sigma_i}\right) + \left(\mathbf{E}\left(\frac{w^i - \mu_i}{\sigma_i}\right)^2 - 1\right) \frac{\pi_1^i}{\sigma_i} \tag{24}$$

$$\frac{\partial \pi_2^i}{\partial \sigma_i} = \Phi\left(\frac{\partial y_2}{\partial \sigma_i}\right) - \Phi\left(\frac{\partial y_1}{\partial \sigma_i}\right) + \left(\mathbf{E}\left(\frac{w^i - \mu_i}{\sigma_i}\right)^2 - 1\right) \frac{\pi_2^i}{\sigma_i} \tag{25}$$

$$\frac{\partial \pi_3^i}{\partial \sigma_i} = \left(\mathbf{E}\left(\frac{w^i - \mu_i}{\sigma_i}\right)^2 - 1\right) \frac{\pi_3^i}{\sigma_i} - \Phi\left(\frac{\partial y_2}{\partial \sigma_i}\right) \tag{26}$$

where $\mathbf{E}(w^i) = \int_{y_{k-1}}^{y_k} w^i \frac{\phi(w^i)}{\pi_k^i} dw^i$, $\mathbf{E}\left(\frac{w^i-\mu_i}{\sigma_i}\right)^2 = \int_{y_{k-1}}^{y_k} \left(\frac{w^i-\mu_i}{\sigma_i}\right)^2 \frac{\phi(w^i)}{\pi_k^i} dw^i$, and $i \in \{p, c\}$. Finally, note that when $\mu_p \neq \mu_c$ and $\sigma^2 = 1$, the unique intersection point is $\frac{\mu_p+\mu_c}{2}$.

Table 2: 1990−1999 log Means and Variances, Growth Rates and Overlap Indices

	Prefecture		County	
	Panel (Restricted) Sample			
Mean: Log Per Capita Incomes (1990,1999)	8.0285	8.6759	7.3368	8.0809
Variance: Log Per Capita Incomes (1990,1999)	0.3027	0.3434	0.1791	0.3310
Average Annual Growth Rates	0.0719		0.0827	
Overlap Index, OI (1990, 1999)	(0.3968 , 0.3898)			
N(0,1) Test for Differences in (Growth Rates, OI)	(2.3279 , -0.2622)			
	Full Sample			
Mean Log Per Capita Incomes (1990,1999)	8.0242	8.5662	7.3325	8.0840
Variance: Log Per Capita Incomes (1990,1999)	0.3030	0.4020	0.1769	0.3192
Average Annual Growth Rates	0.0602		0.0835	
Overlap Index, OI (1990, 1999)	(0.3973, 0.4167)			
N(0,1) Test for Difference in (Growth Rates, OI)	(6.6238, 0.6452)			

Note: For the test of differences, the Z statistic is reported.
Incomes are denominated in Renminbi (RMB).

Table 2 presents the logarithmic means and variances, growth rates, Overlap Indices and Overlap Index Comparisons for County versus Prefecture level cities for 1990 and 1999. The data strongly support the hypothesis that the growth rate in the County level urban income distribution is greater than that of the Prefecture level urban income distribution. Further the results for both the panel and full data set support this view as well. However the full data Overlap Index (based upon integrals of maximum likelihood estimated log normals) does not admit the same inference. It should be noted that both the difference in growth rates test and the convergence test (Based on $\Delta\widehat{OI}$ of equation (7)) on the full sample have accommodated for the between period covariances induced by the partial panel nature of the data (For details of how the accommodation is made see Anderson (2003)). In the case of the panel sample, the Overlap Index records a small decline largely due to the variance in the stochastically dominated County level distribution growing so rapidly, that is implicitly the poorer County level cities are being left behind in the growth race. In the case of the full sample, the Overlap Index records a small but statistically insignificant increase, thus in both cases a null hypothesis of non-convergence could not be rejected. It is interesting to note that while both the means and variances of Prefecture and County level cities are closer together in 1999 than they are in 1990, the degree of likeness in terms of the overlap of the distributions has not changed significantly.

4.2 Example 2: The Effect of Changes in Custody Law on Child Educational Attainment of Single Parent Families

The effect on educational attainment of different types of parental arrangements lay at the heart of the inter-generational income relationship. Leo (2008) studied these effects with respect to single and intact parent families within the context of changes in the custody laws in the United States, and found that family types have a significant impact on a child's educational achievement. Here we illustrate instead the polarizing effect that family type has on the children of single parent families in comparison to children of intact families over three census decades of 1970 to 1990. First, we define children of parents who have either divorced or separated as *endogenously single* and will so refer to them in the rest of the example. A simple model of grade attainment, where a student of age t who started school (grade 1) at age t^* and has a probability p of graduating to the next grade level, predicts that grade attainment in the population of students will have a mean of $1 + p(t - t^*)$ and a variance of $p(1 - p)(t - t^*)$. Assuming p to be a function of family type and different for children of endogenously single and intact parent families, attainment of children of different family types will diverge in the mean but increase in variance with age. The Overlap Index provides an ideal indicator of whether attainments by family type are diverging or converging in a more general sense.

Throughout the 1970s and 1980s, states in the United States continuously changed divorce and custody laws, directly affecting the fortunes of children in *Endogenously Single* parent family situations. A new trend in child custody dispute resolution emerged in the early 1980s in the United States, where previous maternal preference since the 1950s were rescinded in favor of one without bias, coupled with a gradual trend towards statutory leanings toward joint custody awards in custody dispute resolutions. This is exemplified by the fact that before 1980, only 4 states acknowledged joint custody as a possible arrangement in custody awards. However by 1990, only 14 states had not incorporated joint custody. The force of this statutory amendment may be noted from the surge in joint custody awards in California (from 2.2% in 1979 to 13% in 1981 (Maccoby and Mnookin 1994)), and Wisconsin (from 2.2% in 1980-81 to 14.2% in 1991-92 (Brown, Marygold, and Cancian 1997)).

Table 3: Summary Statistics of Grade Attainment by Age and Family Structure

	1970				1980				1990			
	Age 15	Age 16	Age 17	Age 18	Age 15	Age 16	Age 17	Age 18	Age 15	Age 16	Age 17	Age 18
Grade Attainment	Endogenously Single (Proportion in each Attainment Group)											
PreSchool/No Education (1)	0.0083	0.0059	0.0026	0.0057	0.0020	0.0010	0.0016	0.0038	0.0020	0.0011	0.0011	0.0028
Grades 1−4 (2)	0.0031	0.0032	0.0037	0.0057	0.0010	0.0017	0.0022	0.0010	0.0025	0.0000	0.0008	0.0000
Grades 5−8 (3)	0.4974	0.1652	0.0739	0.0474	0.5407	0.1413	0.0526	0.0398	0.4031	0.0912	0.0277	0.0208
Grade 9 (4)	0.4480	0.3743	0.1300	0.0799	0.4329	0.4227	0.1401	0.0775	0.4328	0.3291	0.0884	0.0361
Grade 10 (5)	0.0363	0.4112	0.3485	0.1273	0.0197	0.4038	0.3888	0.1246	0.1468	0.4287	0.3113	0.0930
Grade 11 (6)	0.0031	0.0324	0.3999	0.3904	0.0032	0.0239	0.3747	0.4228	0.0081	0.1381	0.4366	0.3754
Grade 12 (7)	0.0039	0.0078	0.0398	0.3239	0.0005	0.0050	0.0390	0.3156	0.0045	0.0095	0.1274	0.3474
1−3 Years of College (8)	0.0000	0.0000	0.0016	0.0198	0.0000	0.0005	0.0011	0.0149	0.0003	0.0022	0.0068	0.1244
More than 4 Years of College (9)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	3.5258	4.3099	5.1850	5.8833	3.4790	4.3208	5.1975	5.9232	3.7636	4.6479	5.5613	6.3546
Standard Error	0.0139	0.0185	0.0233	0.0325	0.0090	0.0123	0.0156	0.0210	0.0130	0.0148	0.0158	0.0194
Number of Observations	2288	2191	1908	1414	4019	4012	3670	2890	3570	3564	3688	3215
Grade Attainment	Intact Parent Families (Proportion in each Attainment Group)											
Preschool/No Education (1)	0.0046	0.0035	0.0044	0.0051	0.0011	0.0011	0.0013	0.0016	0.0018	0.0008	0.0010	0.0015
Grades 1−4 (2)	0.0035	0.0022	0.0026	0.0042	0.0020	0.0010	0.0015	0.0011	0.0025	0.0000	0.0006	0.0006
Grades 5−8 (3)	0.4366	0.0858	0.0339	0.0264	0.4818	0.0844	0.0242	0.0181	0.3896	0.0556	0.0175	0.0097
Grade 9 (4)	0.5249	0.3684	0.0673	0.0286	0.4947	0.4203	0.0718	0.0322	0.4612	0.3424	0.0472	0.0146
Grade 10 (5)	0.0262	0.5087	0.3515	0.0768	0.0176	0.4698	0.4116	0.0843	0.1344	0.4641	0.3259	0.0556
Grade 11 (6)	0.0028	0.0275	0.5064	0.4229	0.0021	0.0200	0.4557	0.4612	0.0062	0.1273	0.4834	0.4026
Grade 12 (7)	0.0013	0.0035	0.0326	0.4138	0.0007	0.0031	0.0327	0.3869	0.0029	0.0076	0.1194	0.3526
1−3 Years of College (8)	0.0000	0.0003	0.0012	0.0222	0.0000	0.0002	0.0012	0.0147	0.0014	0.0023	0.0050	0.1627
More than 4 Years of College (9)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0002
Mean	3.5784	4.4747	5.4143	6.2029	3.5348	4.4302	5.3950	6.2010	3.7610	4.6928	5.6490	6.5542
Standard Error	0.0037	0.0048	0.0058	0.0081	0.0036	0.0046	0.0053	0.0068	0.0060	0.0067	0.0071	0.0088
Number of Observations	25929	24311	22757	16736	23910	23870	23074	18335	15594	14923	14546	12370
Test Statistic for Δ in Means Test	-3.6481	-8.6441	-9.5271	-9.5365	-5.7345	-8.3455	-11.9630	-12.5960	0.1811	-2.7643	-5.0783	-9.3664

The following analysis draws on data from the Integrated Public Use Microsample Series (IPUMS) of the decennial Census for the decades of 1970 to 1990. The proportion of children of each family type, age group (15 to 18 years of age) and census year, at each level of educational attainment, average (coded) grade attainments[5] and standard deviations, together with the asymptotically normal difference in means test (the null of common "family type" variances is always rejected, so the standard difference in means test is inappropriate) for the years 1970 to 1990 are reported in Table 3. For all three years, as predicted the means and variances grow with age. Further, note the fall in the value of the test statistic for each age group between the 1980s and 1990s cohorts. Superficially, there does seem to be a narrowing of the educational attainment gap over the years. In fact in tests not reported here, it was found that the difference in means across the family types (intact versus endogenously single), narrowed significantly between 1980 and 1990 relative to that between 1970 and 1980.

Table 4: Alienation Indices ($IA = 1 - OV$) and Tests by across Family Structure and Years

	Alientation Index			
Year	Age 15	Age 16	Age 17	Age 18
IA_{1970}	0.0774	0.0978	0.1113	0.1248
	(0.0041)	(0.0045)	(0.0067)	(0.0081)
IA_{1980}	0.0629	0.0662	0.1039	0.1098
	(0.0029)	(0.0029)	(0.0047)	(0.0054)
IA_{1990}	0.0296	0.0487	0.0615	0.0713
	(0.0022)	(0.0033)	(0.0036)	(0.0044)
$IA_{1970} - IA_{1980}$	2.9053	5.9185	0.90478	1.5495
	[0.0018]	[0]	[0.1828]	[0.0606]
$IA_{1980} - IA_{1990}$	9.3524	3.99	7.1264	5.5333
	[0]	[0]	[0]	[0]
$(IA_{1970} - IA_{1980}) -$	-3.086	2.0451	-3.4689	-1.9515
$(IA_{1980} - IA_{1990})$	[0.9990]	[0.0204]	[0.9997]	[0.9745]

Note: Standard errors in parenthesis.
P-values in brackets.

Table 4 provides the alienation ($\widehat{IA} = 1 - \widehat{OI}$) comparisons between the two family structures across the age groups from 1970 and 1990. From the reported Alienation Indices,

[5] The IPUMS codes for educational attainment are as follows: 1 if preschool or no education; 2 if grades 1 to 4; 3 if grades 5 to 8; 4 if grade 9; 5 if grade 10; 6 if grade 11; 7 if grade 12; 8 if 1 to 3 years of college; 9 if more than 4 years of college.

although there is still significant evidence of *alienation*, notice the gradual fall in the Alienation Index across the three decades. Next, examining the differences of the indices across the years, note that the narrowing of the attainment gap is significant across all the adjacent years. Further, notice the significant improvement between 1980 and 1990, compared to the improvement between 1970 and 1980. If we can construe the rate of change between 1970 and 1980 as a trend effect, this change in the Alienation Index between the 1980 and 1990 is suggestive of the possibility that it could have been due to the change in Custodial Laws in the 1980s.

4.3 Example 3: The Effect of Anti-Poverty Policies on Single Parent and Pensioner Households in the United Kingdom

Single Parent and Pensioner households have constituted a significant component of the relative poverty calculation in the United Kingdom and have been targeted sub-populations for public policy. It is significant that poverty level targets were expressed in relative (the proportion of agents experiencing incomes less than some specified proportion of median income) rather than absolute (the proportion of agents experiencing incomes less than some specified proportion of a needs based income measure) terms and reflects the recent[6] popular notion that poverty is a relative concept (see Hills (2001) and Hills (2002))[7]. Indeed measures

[6] This view is not new viz:

> "..By necessaries I understand, not only the commodities which are indispensably necessary for the support of life, but whatever the custom of the country renders it indecent for creditable people, even the lowest order, to be without." Smith (1976)

Similarly Ferguson (1767) states,

> "The necessary of life is a vague and relative term: it is one thing in the opinion of the savage; another in that of the polished citizen: it has a reference to the fancy and to the habits of living."

[7] Reducing the number of children in relative poverty has been a policy target of the British Government since 1998, and the children living in single parent households are a significant component of the calculation. Even after child poverty reductions had been achieved in 2002/3, children of single parents constituted over 40% of poor children (Brewer, Goodman, Myck, Shaw, and Shephard (2004)) but less than a quarter of all children. Concern over pensioners was expressed earlier. Their relative poverty peaked in the late 1980s and declined and stabilized during the late 1990s onwards as the group benefited from improvements in Minimum Income Guarantees and the Basic State Pension. In fact both groups experienced steady declines in absolute poverty throughout the 1990's while trends in relative poverty rates have not been so obvious.

of the relative poverty of subgroups are an expression of how aspects of the subgroup income distribution differ from that of the rest of the population, summarized by the population median. While this falls short of measuring subgroup alienation (differences between the income distributions of a subgroup and its complement), it is very much in that spirit. Indeed if the policy objective is to make the subgroup and its complement more "alike", measures of alienation are the appropriate comparison tool.

Evaluation of the success (or otherwise) of anti-poverty policies in the U.K. has been clouded by controversy over the recent abandonment of some of the comparison instruments, specifically the "after housing cost" income measure (Brewer, Goodman, Myck, Shaw, and Shephard 2004). Poverty measure calculations had been based upon both before and after housing costs income measures, an acknowledgement of the concern that housing expenditures do not reflect economic opportunity costs in the sense that other consumption expenditures do. This issue is particularly pertinent in the case of pensioners and single parents. A large portion of pensioners own their own homes and the nominal expenditures on the property clearly underestimate the welfare gains from inhabiting the property. A large portion of single parents inhabit properties in the social rented sector where the rents have been set with little regard to housing quality or the current market. The presumption has been that the distinction materially effects various welfare and poverty calculations and it would be interesting to compare the individual measures with the consequences of considering the joint impact of after housing cost incomes and housing costs. That is to say, does the distinction alter the extent to which the groups are alienated from the rest of society?

Incomes are calculated from the Family Resources Survey. U.K. Poverty measure calculations are based upon "equivalized" concepts using the McClements (1977) equivalence scale, expressing household incomes as the amount that a childless couple would require to enjoy the same standard of living (see Brewer, Goodman, Myck, Shaw, and Shephard (2004) for details). Incomes before housing costs and incomes net of housing costs are reported as well as housing costs. Table 5 provide summary statistics for single parent and non-single parent households as well as for pensioner and non-pensioner household breakdowns. It is no surprise to learn that single parent and pensioner average and median incomes are less than those of the rest of society (both before and after housing costs). Somewhat more striking is the notion that equivalized housing costs are higher for single parent families than those of the rest of society but lower for pensioners than those of the rest of society. The latter difference is easily rationalized since nominal housing expenses are recorded (rather than imputed rents) and a large portion of senior citizens are owner occupiers with no mortgage obligations. The former difference is somewhat surprising, especially since the differences are strongly statistically significant, perhaps it is related to economies of scale in housing costs. One interesting feature of both pensioner and single parent groups is that the variability of housing costs is relatively stable over time when compared to the rest of the population where the variability of housing costs appears to be growing.

Table 6, panels A and B report alienation indices (where the comparison is made between single versus non-single parent households, and pensioner versus non-pensioner households) based upon income before and after housing cost deductions respectively, while panel C reports alienation indices based upon the joint distribution of after housing cost incomes and housing costs. In the single variate case reported in panels A and B, two alienation indices are calculated, the first is based upon an arbitrary partition of the space into 10 equi-probable intervals over the combined sample and is denoted as IA_N, while the second is based upon estimated intersection points (using the Epanechnikov kernel. For details see Silverman (1986)) denoted as IA_E[8]. In order to examine changes in the degree of alienation, each panel also reports the test for differences against year 2002, reporting both the test statistic and the P-value for the change in IA_N. The close correspondence of IA_N and IA_E is worthy of note. Finally, panel C reports the multivariate equivalents of IA_N together with tests of significance between differences against year 2002 in IA_N[9].

From panels A and B of table 6, it is evident that Single Parent households are more alienated than Pensioner households based on both IA_N and IA_E indices, which in turn implies that income distributions are more unlike the rest of the population both before and after housing costs have been accounted for. It is also evident that Single Parent alienation has been significantly reduced over time (both before and after housing costs) whereas little change of substance has occurred with respect to Pensioners except for the after housing cost index of panel B which indicates insignificant reductions in alienation in the 1996-2002 period together with a significant return to the 1996 level of alienation in 2002. The after housing cost alienation index of panel B is always larger than the before housing cost index

[8]Although IA_N is computationally convenient since it involves the arbitrarily determined partitions, as discussed in the appendix A.1, these arbitrary partitions are the source of potential inconsistency resulting in false null asymptotically. Consequently IA_E, which involves the estimation of the intersection points, is recommended theoretically. However, as shown in Lemma 1 of appendix A.1, and the Monte Carlo study of appendix A.2, this problem is not grave.

[9]In the multivariate case, intersections of the density functions are themselves functions, and its estimation is the subject of further research.

Table 5: Summary Statistics

	Panel A: Household "Equivalized" Incomes (Before Housing Costs)							
	Single Parent Household				Non-Single Parent Household			
Year	Mean	Median	σ	n	Mean	Median	σ	n
1996	186.39	163.43	89.28	2116	298.99	249.13	238.06	28099
1998	212.49	179.51	206.06	1901	330.13	271.67	289.65	25145
2000	240.67	208.65	132.98	2050	368.49	299.85	369.73	25737
2002	254.71	223.79	117.49	1252	378.22	318.76	309.85	16383
	Pensioner Household				Non Pensioner Household			
Year	Mean	Median	σ	n	Mean	Median	σ	n
1996	232.15	192.82	160.62	7216	309.6	261.12	248.05	22999
1998	252.98	207.58	190.38	6541	343.84	286.43	307.33	20505
2000	283.62	235.59	212.49	6823	383.61	315.17	392.27	20964
2002	292.88	252.89	177.99	3733	391.32	332.13	325.81	12518
	Panel B: Household "Equivalized" Incomes (After Housing Costs)							
	Single Parent Household				Non-Single Parent Household			
Year	Mean	Median	σ	n	Mean	Median	σ	n
1996	144.38	114.54	88.59	2098	262.64	219.26	239.22	28240
1998	166.62	128.29	210.37	1880	289.31	238.03	288.83	25295
2000	194.58	152.89	134.42	2030	323.93	264.88	373.07	25982
2002	209.94	169.57	121.48	1238	337.58	286.09	307.84	15126
	Pensioner Household				Non Pensioner Household			
Year	Mean	Median	σ	n	Mean	Median	σ	n
1996	214.02	167.91	170.37	7199	267.15	226.06	248.95	23135
1998	234.3	184.24	201.05	6527	295.48	245.58	305.76	20650
2000	266.27	215	226.43	6807	329.9	271.56	395.37	21213
2002	275.73	229.08	186.71	3720	343.28	295.08	324.02	12644
	Panel C: Household "Equivalized" Housing Costs							
	Single Parent Household				Non-Single Parent Household			
Year	Mean	Median	σ	n	Mean	Median	σ	n
1996	43.69	38.09	33.22	2166	34.72	29.73	39.68	28099
1998	48.15	41.75	37.54	1901	38.69	32.98	45.8	25145
2000	48.44	44.38	35.96	2050	40.79	34.99	51.79	25737
2002	47.64	44.53	33.05	1252	37.16	32.26	45.68	14999
	Pensioner Household				Non Pensioner Household			
Year	Mean	Median	σ	n	Mean	Median	σ	n
1996	18.88	9.46	34.11	7216	40.51	34.19	39.43	22999
1998	19.28	8.81	35.92	6541	45.76	38.6	46.14	20505
2000	18.06	8.66	38.28	6823	48.94	41.48	52.1	20964
2002	18.29	8.17	39.89	3733	43.83	37.91	44.66	14999

Note: "Equivalized" Incomes in denominated £.

Table 6: Alienation Indices and Tests

	Panel A: Household "Equivalized" Income (Before Housing Costs)							
	Single Parent Households				Pensioners			
Year	IA_N	IA_E	Δ over 2002 $(Z, F(Z))$		IA_N	IA_E	Δ over 2002 $(Z, F(Z))$	
1996	0.3602	0.3562	-3.2208	0.0006	0.2423	0.2599	0.1081	0.5430
1998	0.3585	0.3631	-3.0606	0.0010	0.2540	0.2699	-1.0719	0.1419
2000	0.3041	0.3063	0.0107	0.5403	0.2377	0.2553	0.5446	0.7070
2002	0.3043	0.3015			0.2431	0.2597		
	Panel B: Household "Equivalized" Income (After Housing Costs)							
	Single Parent Households				Pensioners			
Year	IA_N	IA_E	Δ over 2002 $(Z, F(Z))$		IA_N	IA_E	Δ over 2002 $(Z, F(Z))$	
1996	0.4075	0.4166	-4.6946	0.0000	0.2453	0.2341	-0.2215	0.4123
1998	0.3949	0.4022	-3.9038	0.0000	0.2301	0.2099	1.3028	0.9037
2000	0.3274	0.3372	-0.1857	0.4264	0.2218	0.2034	2.1657	0.9848
2002	0.3241	0.3322			0.2432	0.2303		
	Panel C: Joint Household "Equivalized" Income (After Housing Costs and Housing Costs)							
	Single Parent Households				Pensioners			
Year	IA_N	IA_E	Δ over 2002 $(Z, F(Z))$		IA_N	IA_E	Δ over 2002 $(Z, F(Z))$	
1996	0.4943		-1.2252	0.1103	0.3337		1.9696	0.9756
1998	0.5013		-1.5921	0.0557	0.3344		1.8748	0.9696
2000	0.4523		1.0160	0.8452	0.3345		1.8779	0.9698
2002	0.4712				0.3554			

Note: "Equivalized" income denominated in £.

of panel A for Single Parents, but this is not the case for pensioners. This is largely the result of the nature of housing costs experienced by the two groups as mentioned before. From table 5, panels A and C, it may be seen that mean (and median) Single Parent and Pensioner before housing cost incomes are always lower than that of the rest of the population, however while equivalized Single Parent housing costs are greater than the rest of the population, Pensioner housing costs are less. Thus when housing costs are removed Single Parent and non-Single Parent income distributions will be further apart and if anything Pensioner and Non-Pensioner Income distributions will be closer together.

A different picture emerges when the joint distributions of after housing costs incomes and housing costs are considered, reported in panel C of table 6. Firstly, alienation is more substantial for both Single Parents and Pensioner households. There is no significant reduction in Single Parent alienation and, at the 5% level, a significant increase in Pensioner Alienation. The latter phenomenon is due to the fact that measured pensioner housing expenses are becoming more unlike the rest of society in that they are moving relatively lower. This is not the case for Single Parent housing costs which, in terms of the median, are becoming relatively higher.

5 Conclusion

The notion of distributional overlap has been exploited in introducing some simple nonparametric and parametric, uni-dimensional and multi-dimensional indices and their tests for considering *alienation* and *convergence* issues, in both discrete and continuous random variable paradigms. The properties of the nonparametric indices, in terms of the biases inherent in the techniques have been assessed in a simple Monte Carlo framework (Their parametric counterparts have more attractive features, in terms of the absence of asymptotic bias, endowed by a parametric structure). Based upon the extent to which two distributions overlap, this paper shows that the indices and tests are easy to implement in a multi-dimensional or multiple characteristic setting. Application of the tests was exemplified in three quite diverse situations: in considering the convergence of the income distributions of different city types in China; in considering the effects of family law reform on the educational attainment of children in single parent families and in comparing multivariate characteristic distributions of family types in the United Kingdom. In each case the indices and tests proved an effective instrument of comparison. Application of these tests need not be confined to the present framework, they could be readily applied wherever there is a need to assess the general degree of commonality or dissimilarity of two distributions. For example they could be used as a specification test of experimental design in the random assignment literature and as an empirical tool in the assortative pairing literatures.

A Appendix

A.1 On the Inconsistency of Point-Wise Comparison Tests

Many tests have been proposed for examining the equality (or otherwise) between two functions over a range by examining their proximity at a sequence of points within that range. Tests based upon the structure of Pearsons Goodness of Fit test including modifications and extensions of it (Anderson (1994); Andrews (1988)), Contingency Table, and Homogeneity of Parallel samples tests, can be interpreted this way (Rao 1973), Lorenz and Generalized Lorenz Dominance tests (Beach and Davidson 1983), Stochastic Dominance Tests in the Poverty, Inequality and Finance Literature (Anderson (1996); Davidson and Duclos (1997); Davidson and Duclos (2000)) are also all members of this class. These tests are often criticized for their potential inconsistency, which results in the failure to reject a false null asymptotically, consequently making the case for less powerful but nonetheless consistent tests (for example Kolmogorov-Smirnov type tests, see Anderson (2001)).

Generally for two smooth and continuous functions $f(x)$ and $g(x)$, defined for $x \in [a, b]$, based upon a random sample (or samples) of a size that grow uniformly with the population size T, the above class of tests can be represented by:

$$P = H(T) \sum_{k=1}^{K+1} \left(\sum_{j=1}^{k} \mathbb{I}(j) \int_{y_{j-1}}^{y_j} (f(x) - g(x))\, dx \right)^2 G_k + O\left(\frac{1}{T}\right) \tag{A-1}$$

where as before y_j, $j = \{1, .., K\}$ represent the K mutually exhaustive ordered partitions (such that $y_1 < y_2 < ... < y_K$), and together with $y_0 = a$ and $y_{K+1} = b$ fully partitions the support $[a, b]$ into $K + 1$ intervals. G_k is a function of the appropriate elements of the inverse of the covariance matrix of integrals of differences that appear in the test and are zero otherwise, and $\mathbb{I}(j)$ is an indicator function. The elements G_k are $O(1)$ asymptotically, and $H(T)$ the sample size factor, is monotonically increasing and at least $O(T)$. For example, in goodness of fit tests, $f(x)$ corresponds to an empirical density function, $g(x)$ to the theoretical density under the null. In the parallel samples tests, $f(x)$ and $g(x)$ are two empirical densities being compared, in both cases $\mathbb{I}(j) = 1$ for $j = k$, 0 otherwise ($H(T) = T$ in the former and $\frac{T_f T_g}{T_f + T_g}$ in the latter where T_f and T_g are the sample sizes from the respective distributions). In First Order Stochastic Dominance (and Lorenz) tests, $f(x)$ and $g(x)$ correspond to empirical density functions (or monotonic transformations of them) with $\mathbb{I}(j) = 1$ for all j. In higher order dominance tests, $f(x)$ and $g(x)$ correspond to higher order integrals (Anderson (1996)) or incomplete moment estimates (Davidson and Duclos 2000) again with $\mathbb{I}(j) = 1$.

Given the sample size factor is $H(T)$ and the covariance factors are $O(1)$, inconsistency of P (when $f(x) \neq g(x)$ except for a finite set of intersection points) requires that:

$$\int_{y_{k-1}}^{y_k} (f(x) - g(x))\, dx = 0, \forall y_k, k = \{1, ..., K+1\} \tag{A-2}$$

The partition points y_k chosen by the investigator are the source of potential inconsistency. In the case of goodness of fit tests, advice abounds as to what and how many y_k's should be chosen (See, for example, Andrews (1988) and Rayner and Best (1989)) but it largely focuses on power issues and ignores the potential inconsistency problem. The following lemma shows that, if the points at which $f(x)$ and $g(x)$ intersect are finite (M) in number, then there are a finite (K) number of partition points that generate the inconsistency property and furthermore $K < M$.

Lemma 1 For smooth, continuous functions $f(x)$ and $g(x)$ defined on $[a, b]$, let there be M ordered interior *intersection points* z_m, such that $f(x) = g(x)$ when $x = z_m$, $m = \{1, 2, .., M\}$ and $f(x) \neq g(x)$ otherwise, except possibly at $x = a$ and $x = b$. Then the *partitions*, y_k $k = \{1, 2, ..., K\}$, satisfying equation (A-2) number at most K where $K < M$.

Proof. Suppose without loss of generality, $f(x) > g(x)$ for $x \in [a, z_1)$, then from the smoothness and continuity assumptions for $f(x)$ and $g(x)$, $|f(x) - g(x)| > 0$ for $x \in (z_m, z_{m+1})$, $m = 1, ..., M-1$. Since:

$$\int_{a}^{y(\leq z_1)} (f(x) - g(x))\, dx > 0 \tag{A-3}$$

there can be no partition point in $[a, z_1)$, otherwise a term of $O(1)$ remains in the extreme left tail of the region $[a, b]$. Similarly, since:

$$(-1)^M \int_{y(\geq z_M)}^{b} (f(x) - g(x))\, dx > 0 \tag{A-4}$$

there can be no partition point in $(z_M, b]$, otherwise a term of $O(1)$ remains at the extreme right tail of the region $[a, b]$. Finally, since:

$$(-1)^m \int_{y_L}^{y_U} (f(x) - g(x))\, dx > 0, \forall y_L < y_U, y_L, y_U \in (z_m, z_{m+1}) \tag{A-5}$$

there can be at most one partition point in (z_m, z_{m+1}) for $m = \{1, 2, ..., M-1\}$, otherwise a term of $O(1)$ remains within the region (z_m, z_{m+1}). Hence there are at most $M-1$ partition points satisfying (A-2). ■

Though the result is "simple", its practical implications are significant. For test inconsistency the set of points satisfying equation (A-2), or a subset of them, have to be chosen exclusively. The number of points, located on an infinite space, has been shown to be finite and bounded from above by the number of intersections of $f(x)$ and $g(x)$ so that, in the assumed circumstances, the probability of choosing them is for all purposes arbitrarily close to zero. When distributions being investigated are uni-modular, the number of intersection points is likely to be small, (unlikely to be more than 4 for example), so that partition schemes need not be extensive for the inconsistency issue to be of no consequence. Multi-modality of the underlying distributions engendered by mixtures will increase the order of the problem slightly but again the degree of multi-modality itself needs to be extensive and the null and alternatives have to be close to present any real prospect of a problem. The result also highlights when inconsistency can arise. If for example $f(x) = g(x)$ over some substantive range of x (as would occur if a policy transferred income from people immediately above some poverty line to people immediately below it whilst leaving the rest of the income distribution unaltered) then an injudicious selection of y_k's, specifically not having a y_k at the poverty line, will engender inconsistency when comparisons are made over the whole distribution. Clearly the y_k's need to be located more intensely within the range over which curves potentially differ. Evidently smoothness and continuity properties are crucial since when distribution functions exhibit substantial mass at a point the potential for inconsistency increases. In short when distributions are smooth and continuous it takes very special circumstances for inconsistency in these tests to arise, either a freakish coincidence or else something that can readily be spotted in advance of testing.

A.2 The Monte Carlo Study

Independent samples were drawn on $f(x) \backsim N(0,1)$ and $g(x) \backsim N(0, 2.25)$. The intersection points y_k, $k \in \{1, 2\}$, for these two distributions is given by $\pm 1.2 \times \sqrt{1.25 \times \ln(2.25)} = \pm 1.2082$ and the exact value of the Alienation Index, IA, for these two distributions is 0.1936. Let $\widehat{f}$ be the kernel density estimator of $f(x)$, and denoting the window width by h, then the bias up to $O(h^2)$ of the kernel estimator $\widehat{f}$ is given in Pagan and Ullah (1999) by:

$$\text{Bias}(\widehat{f}) = \frac{h^2}{2}\mu_2 \frac{d^2 f(x)}{dx^2}$$

where μ_2 is the second moment of the kernel variate and for this case the second derivative of $N(0, \sigma^2)$ is given by:

$$\frac{1}{\sigma^3\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \left(\frac{x^2}{\sigma^2} - 1\right) \tag{A-6}$$

In the region of the intersection in the example, and thus the bias, is positive for $f(.)$ and negative for $g(.)$, which will result in negative bias in the estimate of the lower intersection point and positive bias in the estimate of the upper intersection point.

Any bias in the estimates of the intersection points will always engender negative bias in $\widehat{IA}$ since it always engenders positive bias in the estimate of the Overlap Index, $\widehat{OI}$. To explore this issue three experiments were performed based upon sample sizes of 1000, 5000 and 10000 drawn from each distribution. The intersection points were estimated using Epanechnikov kernel estimates of $f(x) - g(x)$ employing the optimal window width recommended in Silverman (1986) formula 3.31. The magnitude of the increments of x considered was based upon the range of the combined sample divided by 100. The Alienation Index was calculated using both the known values and the estimated values of the intersection points. This exercise was replicated 200 times. Goodness of fit tests for normality based upon 10 equiprobable cells were conducted for both the intersection point estimates and true Alienation Index based upon known and estimated intersection points. The index appears to retain normality in these circumstances as the following table, reporting some simple simulations, attests. The expected bias in the intersection point estimates is apparent, however the bias it engenders in $\widehat{IA}$ is not large because it is swamped by the upward bias due to the conditional probability issue discussed above.

A problem with this approach is that estimation of intersection points becomes precarious in small samples or in the tails of distributions. For the above example when sample sizes were reduced to 500, 200 and 100, more than 2 intersection points were detected 1.5%, 5% and 17.5% of the time respectively. An alternative approach is to note that the index retains it's distributional properties for an arbitrarily defined set of partition points y_k. The $\widehat{IA}$ will be biased downwards since the Overlap Index will be overstated and, when used in a testing environment it does run the risk of substantial power loss and ultimately, given an inopportune choice of partition, of being an inconsistent test. These dangers can be mitigated by choosing a larger number of partition points, y_k, than anticipated intersection points.

Two environments were investigated. The first, where distributions intersect once (location/mean shift), is based upon two distributions $N(a, 1)$ and $N(b, 1)$ where $d = b - a$ was

Table A.1

n = 1000	Mean	σ	Normality Test, $\chi^2(8)$	Pr(Upper Tail)
Lower Intersection Point, y_1	-1.2412	0.1526	7.8	0.4532
Upper Intersection Point, y_2	1.2535	0.1522	3.8	0.8747
IA(Intersection Points known)	0.1916	0.0202	9.6	0.2942
$\widehat{IA}$(Intersection Points estimated)	0.1954	0.0189	8.7	0.3682
n = 5000				
Lower Intersection Point, y_1	-1.2271	0.0738	4.7	0.7891
Upper Intersection Point, y_2	1.2194	0.0709	4.7	0.7891
IA(Intersection Points known)	0.1941	0.009	5.2	0.736
$\widehat{IA}$(Intersection Points estimated)	0.195	0.0094	7.4	0.4942
n = 10000				
Lower Intersection Point, y_1	-1.2206	0.0569	6.0	0.6472
Upper Intersection Point, y_2	1.2188	0.0604	4.3	0.8291
IA(Intersection Points known)	0.194	0.006	7.4	0.4942
$\widehat{IA}$(Intersection Points estimated)	0.1945	0.0058	17.8	0.0228

varied from 0.1 to 1.5 in increments of 0.2. The second, where distributions intersect twice (scale/variance shift), is based upon two distributions $N(0,1)$ and $N(0,1+d)$ where d was varied in the same fashion. The distributions were sampled with sizes (T) ranging from 500 to 2500 in increments of 500 and Alienation Indices ($\widehat{IA}$) calculated based upon K partitions with K set at 5, 10 and 20. Partition points were determined by equi-probable partitioning of the combined sample in one instance and, in order to separate out the two sources of bias, by relocating the nearest partition point to the true intersection point in the second instance. Each experiment was replicated a thousand times and the average value and variance of the index calculated for that experiment. After some data analysis, a parsimonious response surface (Hendry 1983) representation of the bias relationships was specified as:

$$\ln\left(\frac{\widehat{IA}_i}{IA_i}\right) = \beta_0 + \beta_1\frac{1}{T_i} + \beta_2\frac{1}{K_i} + \beta_3\frac{1}{K_i^2} + \beta_4\frac{d_i}{T_i} + \beta_5\frac{1}{K_iT_i} + \beta_6\frac{d_i}{K_i} + \beta_7\ln(AI_i) + \beta_8(\ln(AI_i))^2 + \epsilon_i \quad \text{(A-7)}$$

where T_i represents the sample size, K_i represents the number of partitions, $\widehat{IA}_i$ corresponds to the estimated index, IA_i corresponds to the true value of the index, d_i represents the incremental changes in the respective environments (note that d_i is the location shift variable in table A.2 and scale shift variable in table A.3) listed above and "i" corresponds to the i'th experiment. Since $\widehat{IA} = IA + \text{Bias}$, equation (A-7) can be seen as describing an approximation to the $\frac{\text{Bias}}{IA}$ ratio as a function of the various conditions of the experiment. Bias attributable to ignoring the intersection points can be studied by analysing the change in equation (A-7) when $\widehat{IA}$ is measured ignoring the intersection points, and when $\widehat{IA}$ is measured by incorporating the known intersection points in the partition structure. The three set of parameter estimates are reported in Tables A.2 and A.3 for location shift effects and scale shift effects respectively. Generally the spanned intersection point effect appears to have a much smaller impact than that due to the conditional expectation being smaller than the unconditional expectation. Small sample effects can be assessed by considering $\frac{\beta_1}{T} + \frac{\beta_4 d}{T} + \frac{\beta_5}{KT}$. At the sample means these are -0.0816 and 0.00209 respectively for the conditional expectation induced relative bias and intersection point induced relative bias in the location model, and they are correspondingly 0.0433 and 0.00267 in the scale model.

Table A.2: Mean Shift Effect (One Intersection Point)

Variable	Sample Means	Intersection Points Ignored	Intersection Points Included	Intersection Point Effect
Constant		0.1284	0.1388	-0.0105
		(0.0611)	(0.0658)	(0.0070)
1/Sample Size	0.0009	216.8960	212.5517	4.3443
		(26.0316)	(28.0470)	(2.9750)
1/# Partitions	0.1167	-2.2010	-2.1870	-0.0140
		(0.6532)	(0.7037)	(0.0746)
(1/# Partitions)^2	0.0175	3.1780	3.6792	-0.5012
		(2.3851)	(2.5698)	(0.2726)
Location Shift/Sample Size	0.0007	-158.0895	-155.5545	-2.5350
		(20.5525)	(22.1437)	(2.3488)
(1/# Partitions)(1/Sample Size)	0.0001	-267.2096	-275.8518	8.6422
		(152.5599)	(164.3717)	(17.4351)
Location Shift/# Partitions	0.0933	1.1986	1.1283	0.0703
		(0.1904)	(0.2052)	(0.0218)
ln(IA)	-1.4408	0.1987	0.2097	-0.0110
		(0.0545)	(0.0587)	(0.0062)
(ln(IA))^2	2.7483	0.0965	0.0982	-0.0016
		(0.0114)	(0.0123)	(0.0013)
R^2		0.8940	0.8750	0.4720
σ		0.0610	0.0650	0.0070

†Standard errors are in parenthesis.

Asymptotic bias may be studied by allowing $T \to \infty$, in this case the increase in alienation

Table A.3: Variance Shift Effect

Variable	Sample Means	Intersection Points Ignored	Intersection Points Included	Intersection Point Effect
Constant		0.1024	0.1481	-0.0457
		(0.0827)	(0.0839)	(0.0091)
1/Sample Size	0.0009	357.8138	334.8077	23.0062
		(40.9516)	(41.5672)	(4.5181)
1/# Partitions	0.1167	-4.3911	-3.7243	-0.6669
		(0.7039)	(0.7145)	(0.0777)
(1/# Partitions)^2	0.0175	5.0696	4.6225	0.4471
		(2.3321)	(2.3672)	(0.2573)
Scale Shift/Sample Size	0.0007	-145.1569	-136.9142	-8.2427
		(19.9354)	(20.2351)	(2.1994)
(1/# Partitions)(1/Sample Size)	0.0001	-363.9128	-331.4417	-32.4710
		(149.1667)	(151.4089)	(16.4571)
Scale Shift/# Partitions	0.0933	1.5124	1.2696	0.2428
		(0.1841)	(0.1869)	(0.0203)
ln(IA)	-1.5357	0.1605	0.2152	-0.0547
		(0.0785)	(0.0797)	(0.0087)
(ln(IA))^2	2.8313	0.0880	0.0974	-0.0094
		(0.0168)	(0.0170)	(0.0018)
R^2		0.8640	0.8450	0.8010
σ		0.0590	0.0600	0.0070

†Standard errors are in parenthesis.

brought about by increases in location and/or scale parameters increases the relative bias due to conditional expectation effect, but reduces the negative bias due to the intersection point effect, and increasing the number of partitions dilutes the impact of increased alienation on the biases. The marginal effect of increasing the number of partitions is given by $-\frac{\beta_2}{K^2} - \frac{\beta_3}{K^3} - \frac{\beta_6 d}{K^2}$, assuming $d = 1$ and $K = 10$ it can be seen that the effect is positive for both location and scale problems and is positive for both the conditional expectation effect and the intersection point effect[10]. The latter phenomenon is not surprising, smaller intervals generally imply smaller approximation biases at the intersection points. If K is also allowed to go to infinity (but at a slower rate than T) then we observe from the significance of the intercept and ln(IA) terms (the only ones that remain as $\frac{1}{T}$ and $\frac{1}{K} \to 0$) that relative bias

[10]The calculations in the location equations are 0.003228 and 0.000335 for the conditional expectation effect and intersection point effect respectively, and for the scale equations they are respectively 0.02368 and 0.003347.

is present asymptotically. The absolute magnitude of the bias may be obtained by noticing that Bias $= \left[\exp\left(\ln\left(\frac{\widehat{IA}}{IA}\right)\right) - 1\right] \times IA$ and these biases are reported in Table A.4. As may be seen the total bias is always positive and "U" shaped since the bias due to conditional rather than marginal probabilities swamps the intersection bias. Naturally the magnitude of intersection related bias increases with the number of intersections.

Table A.4: Approximate Absolute Asymptotic Bias (Large K)

IA	Total Bias (Location Shift)	Total Bias (Scale Shift)	Intersection Bias (Location Shift)	Intersection Bias (Scale Shift)
0.1	0.0183	0.0199	0.0006	0.0030
0.2	0.0117	0.0144	0.0006	0.0036
0.3	0.0087	0.0110	0.0001	0.0020
0.4	0.0109	0.0117	-0.0007	-0.0014
0.5	0.0185	0.0167	-0.0018	-0.0062
0.6	0.0312	0.0260	-0.0032	-0.0121
0.7	0.0489	0.0394	-0.0047	-0.0192
0.8	0.0711	0.0568	-0.0065	-0.0272
0.9	0.0977	0.0778	-0.0084	-0.0360

References

Akerlof, G. (1997): "Social Distance and Social Decisions," *Econometrica*, 65(5), 1005–1027.

Anderson, G. (1994): "Simple Tests of Distributional Form," *Journal of Econometrics*, 62, 265–276.

——— (1996): "Nonparametric Tests for Stochastic Dominance in Income Distributions," *Econometrica*, 64, 1183–1193.

——— (2001): "The Power and Size of Nonparametric Tests for Common Distributional Characteristics," *Econometric Reviews*, 20, 1–30.

——— (2003): "Poverty in America 1970-1990: Who Did Gain Ground?," *Journal of Applied Econometrics*, 18, 621–640.

——— (2004a): "Making Inferences about the Welfare Poverty and Polarization of Nations: A Study of 101 Countries 1970-1995," *Journal of Applied Econometrics*, 19, 537–550.

——— (2004b): "Towards and Empirical Analysis of Polarization," *Journal of Econometrics*, 122, 1–26.

——— (2005): "Polarization," Entry in the New Palgrave Dictionary of Economics, forthcoming.

Anderson, G., and Y. Ge (2009): "A New Approach to Convergence: The "Complete" Convergence of Post Reform Chinese Urban Income Distributions," Forthcoming in the *Journal of Income Distribution*.

Andrews, D. (1988): "Chi-Square Diagnostic Tests for Econometric Models: Theory," *Econometrica*, 6, 1419–1453.

Atkinson, A. B. (1998): *Social Exclusion, Poverty and Unemployment*In *Exclusion, Poverty and Opportunity*, edited by A. B. Atkinson and J. Hills. Centre for the Analysis of Social Exclusion.

Barrett, G., and S. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

Beach, C., and R. Davidson (1983): "Unrestricted Statistical Inference with Lorenz Curves and Income Shares," *Review of Economic Studies*, 50, 723–735.

Brewer, M., A. Goodman, M. Myck, J. Shaw, and A. Shephard (2004): "Poverty and Inequality in Britain: 2004," *The Institute For Fiscal Studies Commentary 96*.

Brown, P., M. Marygold, and M. Cancian (1997): "Physical Custody in Wisconsin Divorce Cases, 1980-1992," Institute for Research on Poverty, University of Wisconsin-Madison, Discussion Paper no. 1133-97.

Champernowne, D. G. (1953): "A Model of Income Distribution," *Economic Journal*, 83, 318–351.

Davidson, R., and J.-Y. Duclos (1997): "Statistical Inference for the Measurement of the Incidence of Taxes and Transfers," *Econometrica*, 65, 1453–1466.

——— (2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–1464.

Deutsch, J., and J. Silber (1997): "Gini's "Transvariazione" and the Measurement of Distance Between Distributions," *Empirical Economics*, 22, 547–554.

Dhrymes, P. (1970): *Econometrics, Statistical Foundations and Applications*. Harper International.

Duclos, J.-Y., J. Esteban, and D. Ray (2004): "Polarization: Concepts, Measurement, Estimation," *Econometrica*, 72, 1737–1772.

Esteban, J., and D. Ray (1994): "On the Measurement of Polarization," *Econometrica*, 62, 819–852.

Gibrat, R. (1930): "Une Loi Des Repartitions Economiques: Leffet Proportionelle," *Bulletin de la Statistique generale de la France et du Service d'observation des prix*, July-September, 469–451.

Hendry, D. (1983): "Monte Carlo Experimentation in Econometrics," In *Handbook of Econometrics*, Volume 2, edited by Z. Griliches and M. Intriligator. North Holland, Amsterdam.

Hills, J. (2001): "Poverty and Social Security: What Rights? Whose Responsibilities?," In *British Social Attitudes: The 18th Report-Public Policy, Social Ties*, edited by A. Park, J. Curtice, K. Thomson, L. Jarvis and C. Bromley. London: Sage.

——— (2002): "Following or Leading Public Opinion? Social Security Policy and Public Attitudes Since 1997," *Fiscal Studies*, 23, 539–558.

Leo, T. W. (2008): "From Maternal Preference to Joint Custody: The Impact of Changes in Custody Law on Child Educational Attainment," mimeo University of Toronto, Department of Economics.

Maccoby, E., and R. Mnookin (1994): *Dividing the Child: Social and Legal Dilemmas of Custody*. Harvard University Press.

McClements, L. (1977): "Equivalence Scales for Children," *Journal of Public Economics*, 8, 191–210.

Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics*. Cambridge University Press.

Pareto, V. (1897): *Cours dEconomie Politique*. Rouge et Cie, Paris.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications*. John Wiley & Sons.

Rayner, J., and D. Best (1989): *Smooth Tests of Goodness of Fit*. Oxford University Press.

Reed, W. (2001): "The Pareto, Zipf and Other Power Laws," *Economics Letters*, 74, 15–19.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability. Chapman and Hall.

Smith, A. (1976): *An Enquiry Into the Nature and Causes of the Wealth of Nations*. Liberty Classics.

Stine, R., and J. Heyse (2001): "Non-parametric Estimates of Overlap," *Statistics in Medicine*, 20, 215–236.

Weitzman, M. (1970): "Measures of Overlap of Income Distributions of White and Negro Families in the U.S.," Technical Paper 22 Bureau of the Census.