Distributional Overlap: Simple, Multivariate, Parametric and Non-Parametric Tests for Alienation, Convergence and general distributional difference issues.

Gordon Anderson (Univeristy of Toronto)
Ying Ge (University of International Business and Economics, Beijing)
Teng Wah Leo (University of Toronto)

Department of Economics
University of Toronto

June 2005

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.

Introduction.

In keeping with the Akerlov (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 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.

Duclos et al (2004) and Esteban and Ray (1994) provided indices which identify the phenomenon in the mixture of group distributions in a uni-variate framework. Anderson (2004, 2004a), in studying the anatomy of polarized states, provided tests (again in a uni-variate 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 formal definition of alienation in that two identical within group economic variable distributions will yield 0 alienation in the refined definition but will yield alienation measures that increase with the variance of the distribution in the formal case[1].

Weitzman (1970) first proposed a non - parametric version of the overlap measure[2] for the one dimensional case where pdf'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^t may be written as:

$$OI^t = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots \min(f(x,y,z,..), g(x,y,z,..))\,dx\,dy\,dz..$$

Obviously OI^t is bounded between 0 and 1 corresponding to no overlap and "complete" overlap respectively, consequently 1 - OI provides an index of the degree to which two populations are polarized or alienated. Here discussion will continue in terms of the uni-variate case though examples of multivariate situation will be presented. In the following the parametric overlap index and its distribution is outlined in Section 1, Section 2 discusses issues concerning the implementation of the non-parametric version, three examples illustrating uses of the index are proffered in section 3 and section 4 concludes.

1. The Parametric Overlap index

Given consistent and asymptotically efficient estimates of θ_f and θ_g, the respective J x 1 parameter vectors of f( ) and g( ), let q_{fk} be the k'th element of a (K+1) x 1 vector of estimates of

---

[1]Alienation tests have long existed in the econometrics literature, see for example (Dhrymes(1970)). Typically, given co-varying vectors y and x of dimension m x 1 and n x 1 respectively with a conformably partitioned co-variance matrix Σ of the form:

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

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

[2]The Weitzman (1970) overlap measure is of the form $[\int_{\min(x|f(x))}^{x^*} f(x)dx + \int_{x^*}^{\max(x|g(x))} g(x)dx\,]$ where x* is the unique point at which f(x) and g(x) intersect.

the probability of falling in the k'th of the K+1 mutually exclusive and exhaustive intervals on the range of x under f( ) defined by K partition points, where π_{fk} is the true probability. Then, following Rao (1973) p391, d_f, a vector whose typical k'th element is $n^{0.5}((q_{fk}-\pi_{fk})/\pi_{fk}^{0.5})$ is distributed as:

$$d_f \overset{a}{\sim} N(0,(M_f H_f^{-1} M_f' \Omega_f M_f H_f^{-1} M_f'))$$

where $M_f'M_f = H_f$ where M_f is a K x J matrix whose typical kj'th element is given by:

$$M_{kj} = \frac{1}{\sqrt{\pi_k}} \cdot \frac{\partial \pi_k}{\partial \theta_j}$$

where the rank of M is J and where Ω is given by the matrix:

$$\Omega = \begin{bmatrix} \pi_1 & 0 & . & 0 \\ 0 & \pi_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \pi_k \end{bmatrix} - \begin{bmatrix} \pi_1 \\ \pi_2 \\ . \\ \pi_k \end{bmatrix} [\pi_1 \quad \pi_2 \quad . \quad \pi_k]$$

Note that a vector d_g can be similarly defined under g( ) for the same set of K points. Letting q_{min} be formed from the minimum values of the corresponding elements of q_f and q_g and let π_{min} be similarly formed from π_f and π_g and let Ω_{min} also be similarly formed from Ω_f and Ω_g, then a correspondingly defined d_{min} will be distributed as:

$$d_{min} \overset{a}{\sim} N(0,(M_{min} H_{min}^{-1} M_{min}' \Omega_{min} M_{min} H_{min}^{-1} M_{min}'))$$

so that $n^{0.5}(q_{min}-\pi_{min})$ is distributed as:

$$\sqrt{n}(q_{min} - \pi_{min}) \overset{a}{\sim} N(0,(BM_{min} H_{min}^{-1} M_{min}' \Omega_{min} M_{min} H_{min}^{-1} M_{min}' B'))$$

where B is a diagonal matrix with $\sqrt{\pi_{min,k}}$ as the k'th element on its diagonal.

Define an estimate of OI^t as $OI = i'q_{min}$ where i is a (K+1) x 1 unit vector so that:

$$OI \overset{a}{\sim} N(i'\pi_{\min}, i'((BM_{\min}H_{\min}^{-1}M_{\min}'\Omega_{\min}M_{\min}H_{\min}^{-1}M_{\min}'B'))i/n)$$

Clearly as $K \to \infty$ (but $K/n \to 0$), $OI \to OI^t$, alternatively if the K partition points are defined by the intersection points of f( ) and g( ) which, given consistent estimates of the θ's, can be consistently estimated by the solution or solutions to $f(x, \theta_f) = g(x, \theta_g)$ then $OI \to OI^t$ as $n \to \infty$.

For independently sampled observations in states 1 and 2, changes in the degree of convergence can be examined by focussing on $DOI = OI_1 - OI_2$ which, under the null of no change and with the respective variances defined above written as V_1 and V_2, will be distributed as:

$$DOI \overset{a}{\sim} N(0, (V_1 + V_2))$$

2.The Non-Parametric Overlap Index.

For the single characteristic case, for K known or predetermined intersection points x_k , k=1,..,K[3] this may be implemented empirically by the sample sums of:

$$\max[\frac{(I(x-x_K)}{n_x}, \frac{I(y-x_K)}{n_y}] - \sum_{k=1}^{K}\min[\frac{(I(x-x_k)-(I(x-x_{k-1}))}{n_x}, \frac{(I(y-x_k)-(I(y-x_{k-1}))}{n_y}]$$

where x corresponds to the variable associated with f() and y corresponds to the variable associated with g(), I(z) is an indicator function where I(z) = 1 if $z \le 0$ and is 0 otherwise and x_0 is chosen such that $I(x-x_0)=I(y-x_0) = 0$. When the samples on x and y are independent the asymptotic

[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 arrises when too few partitions are chosen as Appendix 1 demonstrates.

distribution of the index is well defined. Letting q_x be a (K+1) x 1 vector whose typical element is given by the proportion of sample elements falling in the K+1 intervals defined by the intersection points x_k then:

$$q_x \overset{a}{\sim} N(\pi_f, (diag(\pi_f) - \pi_f \pi_f{}') / n_x)$$

where π_f is a K+1 vector of the probabilities of falling in the K+1 intervals defined by the intersection points x_k under f( ). The vector q_y can be defined in a similar fashion under g( ) for the same set of points x_k , k=1,..,K so that:

$$q_y \overset{a}{\sim} N(\pi_g, (diag(\pi_g) - \pi_g \pi_g{}') / n_y)$$

When g() is the stochastically dominated distribution, the vector q_x - q_y will alternate in sign with the first element being negative. Letting q_{min} be formed from the minimum values of the corresponding elements of q_x and q_y and let π_{min} be similarly formed from π_f and π_g then IOL = 1-i'q_{min} where i is a (K+1) x 1 unit vector so that:

$$q_{\min} \overset{a}{\sim} N(\pi_{\min}, (diag(\pi_{\min}) - \pi_{\min}\pi_{\min}{}') / n^*)$$

where n* alternates with the appropriate sample size deflator n_x or n_y so that:

$$IA \overset{a}{\sim} N(1 - i'\pi_{\min}, i'((diag(\pi_{\min}) - \pi_{\min}\pi_{\min}{}') / n^*)i)$$

For independently sampled observations in states 1 and 2, changes in the degree of alienation can be examined by focussing on DIA = IA_1 - IA_2 which, under the null of no change in alienation, will be distributed as :

$$DIA \overset{a}{\sim} N(0, i'((diag(\pi_{\min 1}) - \pi_{\min 1}\pi_{\min 1}{}') / n^{*1})i + i'((diag(\pi_{\min 2}) - \pi_{\min 2}\pi_{\min 2}{}') / n^{*2})i)$$

In practice the non-parametric multi-characteristic case will soon run into the curse of dimensionality that bedevils non-parametric estimation, however its implementation simply demands that the vectors q_f, q_g, q_{min} and π_f, π_g and π_{min} correspond to the vectorized list of cell proportions and probabilities generated by partitions of the support of x, y, z,... 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.

Practical Considerations in Implementing the Non-Parametric Test.

In parametric versions of the index estimates of the intersection points and the index are at least asymptotically unbiased, but in the non-parametric versions they are not. There are two sources of bias in the estimate of the "non-parametric" overlap measure which work in opposite directions. One source is the bias induced by estimating the intersection points[4] and is associated with intervals which span the true intersection points and exaggerates the degree of overlap (hence understating the degree of alienation). It is best illustrated by considering f(x) and g(x) defined on the interval [a, b] where $g(x^*) = f(x^*)$ for some $x^* \in (a, b)$ with $g(x) < f(x)$ for $x < x^*$ and $g(x) > f(x)$ for $x > x^*$. The component appropriate for the overlap measure is given by:

$$A = \int_a^{x^*} g(x)dx + \int_{x^*}^{b} f(x)dx$$

whereas the component included in the calculation will be the smaller of either:

$$\int_a^b f(x)dx = A + \int_a^{x^*} (f(x) - g(x))dx$$

[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 + δ x*, the point at which the functions intersect can be estimated by:

$x^* = (|f(x + \delta)-g(x + \delta)|*x+|f(x)-g(x)|*(x + \delta))/(|f(x)-g(x)|+|f(x + \delta)-g(x + \delta)|)$

or:

$$\int_{a}^{b} g(x)dx = A + \int_{x^*}^{b} (g(x) - f(x))dx$$

In both cases the second term on the right hand side is positive hence the degree of overlap will be exaggerated.

The other source of bias is related to the Min(p,q) function, where p and q are the respective independent estimates of the probability of being in an interval under f( ) and g( ), 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 p and q, $E(p) \geq E(p|\, p<q)$. Since p is an unbiased estimator for f() the conditional estimator implicit in the Min function will be downward biassed. Generally for independent p and q, with respective pdf's j(p) and k(q), $E(p|p<q)$ is given by:

$$\int_{0}^{1}\int_{p}^{1} pj(p)k(q)dq\,dp = \int_{0}^{1} pj(p)[1-K(p)]dp = E(p) - \int_{0}^{1} pj(p)K(p)dp$$

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

3. Some Examples.

Three examples are reported exemplifying the use of the overlap measure. The first demonstrates its use in a multivariate non-parametric environment in considering the plight of single parent families and pensioners in the U.K. The second provides a uni-variate parametric application to the issue of divergent or convergent economic growth in Chinese cities with differing administrative structures. The third is a uni-variate non-parametric discrete distribution example studying the impact of different types of parenting arrangements and the grade attainments of young people.

Example 1. Single Parent and Pensioner households have constituted a significant components of the relative poverty calculation in the U.K. and have been targeted sub-populations . 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[5] popular notion that poverty is a relative concept (see Hills (2001, 2002))[6]. Indeed measures 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

---

[5] 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." Adam Smith (1776). 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".

[6]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 et. Al. (2004)) but less than a quarter of all children. Concern over pensioners was expressed earlier. Thier relative poverty peaked in the late 1980's and declined and stabilized during the late 1990's 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.

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 et. Al. (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 under-estimate 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. UK 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 et. Al. (2004) for details). Incomes before housing costs and incomes net of housing costs are reported as well as housing costs. Tables 1 through 3a 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

Table 1. Household “Equivalized” Incomes (Before Housing Costs) Summary Statistics

	Single Parent Household				Non Single Parent Household			
	Mean	Median	Std Dev	n	Mean	Median	Std Dev	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

Table 1a. Household “Equivalized” Incomes (Before Housing Costs) Summary Statistics

	Pensioner Household				Non Pensioner Household			
	Mean	Median	Std Dev	n	Mean	Median	Std Dev	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

Table 2. Household “Equivalized” Incomes (After Housing Costs) Summary Statistics

	Single Parent Household				Non Single Parent Household			
	Mean	Median	Std Dev	n	Mean	Median	Std Dev	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

Table 2a. Household “Equivalized” Incomes (After Housing Costs) Summary Statistics

	Pensioner Household				Non Pensioner Household			
	Mean	Median	Std Dev	n	Mean	Median	Std Dev	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

Table 3. Household "Equivalized" Housing Costs Summary Statistics

	Single Parent Household				Non Single Parent Household			
	Mean	Median	Std Dev	n	Mean	Median	Std Dev	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

Table 3a. Household "Equivalized" Housing Costs Summary Statistics

	Pensioner Household				Non Pensioner Household			
	Mean	Median	Std Dev	n	Mean	Median	Std Dev	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

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 the nominal housing expenses 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.

Tables 4 and 5 report alienation indices based upon income before and after housing cost deductions and table 6 reports alienation indices based upon the joint distribution of after housing cost incomes and housing costs. In the single variate case alienation indices based upon an arbitrary partition of the space based upon 10 equi-probable intervals over the combined sample (IA(NI)) and indices based upon estimated intersection points (IA(IE)) (using the Epanechinikov kernel see Silverman (1983)) are reported together with tests for differences in IA(NI) between years. Table 6 reports the multivariate equivalents of IA(NI) together with tests of significance between year differences in IA(NA). The close correspondence of IA(NI) and IA(IE) is worthy of note.

It is evident that Single Parent households are more alienated than Pensioner households in the sense that their 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 which indicates insignificant reductions in alienation in the 1996-2000 period together with a significant return to the 1996 level of alienation in 2002. The after housing cost alienation index is always larger than the before housing cost index 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. From tables 1, 1a, 3 and 3a 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. 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.

Table 4. Household "Equivalized" Income Alienation Indices and Tests (Before Housing Costs).

	Single Parent Households			Pensioners		
Year	IA(NI)	IA(IE)	Change over 2002 (Z,F(Z))	IA(NI)	IA(IE)	Change 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	

Table 5. Household "Equivalized" Income Alienation Indices and Tests (After Housing Costs).

	Single Parent Households			Pensioners		
Year	IA(NI)	IA(IE)	Change over 2002 (Z,F(Z))	IA(NI)	IA(IE)	Change 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	

Table 6. Joint Household "Equivalized" Income Alienation Indices and Tests (After Housing Costs and Housing Costs).

	Single Parent Households		Pensioners	
Year	IA(NI)	Change over 2002 (Z,F(Z))	IA(NI)	Change 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	

Example 2. In 1978 China embarked upon a series of Economic Reforms which have had a profound impact on the chinese economy. In particular the reforms have had a considerable impact on the urban economy and structure. The unique administrative hierarchical structure of the Chinese Urban System was constructed in the command economy period of the 1950's 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 prefectural cities, and the growth of investment in the state manufacturing sector was the main determinant of urban income growth. 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 helping reduce the "power" gap between the cities with different administrative levels. Secondly economic liberalization, through stimulating rapid private sector growth and shrinking the relative size of the state sector, favoured the county level cities which were characterized by relatively larger private sectors 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 is did these changes result in the urban income size distributions of these two city types becoming more alike. Following (Anderson and Ge (2004)) 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), Pareto and Double Pareto distributions also have a claim as candidates (Pareto(1897), Champernowne(1953) and Reed(2001)), Table 7 reports goodness of fit tests and upper tail probabilities for the comparisons. The evidence for the log normal distribution is compelling 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%.

Table 7.

Goodness of Fit Tests and Upper Tail probabilities in ( ) for Log-Normal and Pareto and Double Pareto Distributions (Prefecture and County level sample sizes respectively in [ ])

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

Calculating the Overlap Index requires calculation of the intersection points for two log normals which was based upon the formula:

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

where, for two normal distributions A and B with respective means μ_A and μ_B and respective standard deviations σ_A and σ_B where $\sigma_A > \sigma_B$ $\delta = \mu_A - \mu_B$ and $\sigma^2 = (\sigma_A / \sigma_B)^2$. Note that for $\mu_A \neq \mu_B$ and $\sigma^2 = 1$ there will be only one intersection point. Formula for the elements of the matrix $\partial\pi/\partial\theta$ where, in the case of the normal distribution, $\theta = (\mu,\sigma)$ are given by $\partial\pi/\partial\mu = (E_{ab}(X) -$

$\mu)\pi/\sigma$ and $\partial\pi/\partial\sigma\mu = (E_{ab}((X-\mu)/\sigma) - 1)\pi/\sigma$ where $\pi = P(a<X<b)$ and $E_{ab}(X)$ is the expectation of X taken over the interval (a, b), expressions for these expectations are available in Johnson et. al. (1994).

Table 8 presents the logarithmic means and variances, growth rates, overlap measures and Convergence 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, results for both the panel and the full data set support this view. However the Complete Convergence index (based upon integrals of maximum likelihood estimated log normals) does not admit the same inference. It

Table 8. 1990-1999 Logarithmic means and variances, Growth Rates overlap measures and Convergence Index Comparisons, full and restricted (i.e. 1990 panel based) samples.

	Prefecture	County
Panel Sample Growth Rate and Convergence Comparisons		
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
Convergence Indices (1990, 1999)	(0.3968, 0.3898)	
N(0,1) Test for Differences in (Growth Rates, CI)	(2.3279, -0.2622)	
Full Sample Growth Rate and Convergence Comparisons		
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
Convergence Indices (1990, 1999)	(0.3973, 0.4167)	
N(0,1) Test for Difference in (Growth Rates, CI)	(6.6238, 0.6452)	

should be noted that both the difference in growth rates test and the complete convergence test have accommodated the between period co-variances 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 index records a small decline largely due to the variance in the stochastically dominated county level distribution growing so rapidly, implicitly the poorer county level cities are being left behind in the growth race. In the case of the full sample, the convergence index records a small but statistically insignificant increase, thus in both cases a Null 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.

Example 3. The effect on educational attainment of different types of parental arrangements lays at the heart of the inter-generational income relationship. Leo (2005) has studied these effects with respect to single parent and two parent families and changes in the custody laws in the United States and finds that the type of family has a significant impact on a child's educational achievement (quality). Here we illustrate the polarizing effect that family type has on the children of single parent families. Two fundamental single parent situations are identified, Exogenously Single, wherein the parent is widowed and Endogenously Single, wherein the parent is divorced or separated. A simple model of grade attainment, whereby a student starts 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 endogenously single and exogenously single parent families, attainment of children of different family types will diverge in the mean but increase in variance with age. An overlap measure provides an ideal indicator of whether attainments by family type are diverging or converging in a more general sense.

Throughout the 80's states continuously changed divorced laws to allow for more joint custody arrangements a side effect of which was to improve the lot of children in Endogenously single parent family situations. A new trend in child custody dispute resolution emerged in the early 1980s in the U.S., where previous maternal preference since the 1950s were rescinded in favor of one without any sex based 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 has 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 & Mnookin, 1994)), and Wisconsin (from 2.2% in 1980-81 to 14.2% in 1991-92 (Brown et. al., 1997)).

The average grade attainments and their standard deviations for the children of endogenously single and exogenously single parents 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 year 1980 are reported in Table 9. As predicted the means and variances grow with the cohort age though the exogenous and endogenous means are only significant at the 5% level for 16 and 18 year old cohorts.

Table 9 (1980)

	Age 15	Age 16	Age 17	Age 18
Endogenous				
Sample Size	3266	3268	2982	2355
Mean Grade	3.4939	4.3433	5.2290	5.9567
Std Dev	0.01018	0.01366	0.01725	0.02350
Exogenous				
Sample Size	976	1122	1183	1074
Mean Grade	3.5225	4.3886	5.2688	6.0326
Std Dev.	0.01883	0.02319	0.02707	0.03340
Difference in Means Test	-1.3391	-1.6817	-1.2390	-1.8584

Table 10 [1- OI] and Dif-Dif Alienation Measures and Tests (1980)

	Age 16 [0.03233]	Age 17 [0.05958]	Age 18 [0.05042]
Age 15 [0.02836]	0.4616 (0.3222) 0.4856 (0.6864)	3.0789 (0.0010) 0.2903 (0.6142)	2.1894 (0.0143) 1.0258 (0.8475)
Age 16 [0.03233]		2.6766 (0.0037) -0.1313 (0.4478)	1.7882 (0.0369) 0.6256 (0.7342)
Age 17 [0.05958]			-0.7995 (0.7880) 0.6950 (0.7565)

Table 10 provides the alienation (i.e. 1-OI) comparisons between the cohorts for 1980 for comparison the asymptotically normal difference in mean differences (Dif-Dif) comparisons are also provided. As may be observed the Dif-Dif test is never significant whereas the overlap index indicates significant alienation at the 5% level with cohort age for the 15-17, 15-18, 16-17 and 16-18 comparisons. Tables 11 and 12 present the 1990 results corresponding to Tables 9 and 10. Again means and variances grow with age cohort but note that the difference in means test is now always of the opposite sign and is significant in cohorts 16 and 17. However as Table 12 reports none of the Dif-Dif or Alienation tests reject the null of no alienation thus supporting the view that custodial law reform has had an impact via the spread of joint custody arrangements in endogenous single parent situations.

Table 11 (1990)

	Age 15	Age 16	Age 17	Age 18
Endogenous				
Sample Size	4365	4360	4433	3806
Mean Grade	3.7033	4.5732	5.4805	6.2709
Std Dev	0.01134	0.01335	0.01468	0.01808
Exogenous				
Sample Size	932	961	1052	983
Mean Grade	3.6652	4.5047	5.3964	6.2177
Std Dev.	0.02580	0.02938	0.03067	0.03598
Difference in Means Test	1.3515	2.1223	2.4736	1.3210

Table 12. [1- Overlap] Alienation Dif-Dif Measures and Tests (P-values)

	Age 16 [0.04293]	Age 17 [0.03729]	Age 18 [0.04563]
Age 15 [0.03274]	1.0538 (0.1460) -0.7095 (0.2390)	0.4970 (0.3096) -1.0416 (0.1488)	1.3081 (0.0954) -0.3072 (0.3793)
Age 16 [0.04293]		-0.5810 (0.7194) -0.3328 (0.3697)	0.2596 (0.3976) 0.2965 (0.6166)
Age 17 [0.03729]			0.8428 (0.1997) 0.5863 (0.7212)

4. Conclusions.

The notion of distributional overlap has been exploited in introducing some simple non-parametric and parametric uni and multi-dimensional tests and indices for considering alienation and convergence issues in both discrete and continuous random variable paradigm's. The properties of the non-parametric 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, the indices and tests have been shown to be easy to implement in a multi-dimensional or multiple characteristic setting. Application of the tests was exemplified in three quite diverse situations, in comparing multivariate characteristic distributions of family types in the U.K., in considering the convergence of the income distributions of different city types in China and in considering the effects of family law reform on the educational attainment of children in single parent families. In each case the indices an 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.

References.

Akerlov G. (1997) “Social Distance and Social Decision” Econometrica 65 1005-1027.

Anderson G.J. (1994) “Simple Tests of Distributional Form” Journal of Econometrics 62, 265-276.

Anderson G.J., (1996) "Nonparametric Tests for Stochastic Dominance in Income Distributions" Econometrica 64, 1183-1193.

Anderson G.J. (2001) “The Power and Size of Nonparametric Tests for Common Distributional Characteristics” Econometric Reviews 20, 1-30.

Anderson G.J. (2003) “Poverty in America 1970-1990: Who Did Gain Ground?” Journal of Applied Econometrics. 18 621-640.

Anderson G.J. (2004) Making Inferences about the Welfare Poverty and Polarization of Nations: A Study of 101 Countries 1970-1995. Journal of Applied Econometrics 19 537-550.

Anderson G.J. (2004a) “Towards and Empirical Analysis of Polarization” Journal of Econometrics 122 1-26.

Anderson G.J. (2005) “Polarization” Entry in the New Palgrave Dictionary of Economics forthcoming.

Anderson G.J. and Y. Ge (2004) “A New Approach to Convergence: The “Complete” Convergence of Post Reform Chinese Urban Income Distributions”. Mimeo University of Toronto Economics Department.

Andrews D.W.K. (1988) “Chi-Square Diagnostic Tests for Econometric Models: Theory Econometrica 6, 1419-1453.

Atkinson A.B. (1998) “Social Exclusion, Poverty and Unemployment” chapter 1 in Exclusion, Poverty and Opportunity A.B. Atkinson and J. Hills eds. Centre for the Analysis of Social Exclusion.

Barrett G and Donald S. “Consistent Tests for Stochastic Dominance” Econometrica 71 71-104, 2003.

Beach, C.M. 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. (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.

Davidson R. And J-Y Duclos (2000) “Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality” Econometrica 68 1435-1464.

Dhrymes P.J. (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.

Ferguson A. (1767) A History of Civil Society. Edinburgh.

Gibrat, R. (1930) “Une Loi Des Repartitions Economiques: L’effet Proportionelle” Bulletin de Statistique General, France, 19, 469.

Hendry D.F. (1983) “Monte Carlo Experimentation in Econometrics” Handbook of Econometrics volume 2 Z. Griliches and M.D. Intriligator (editors), North-Holland

Hills, J. (2001) “Poverty and Social Security: What Rights? Whose Responsibilities?” in A. Park, J. Curtice, K. Thompson, L. Jarvis and C. Bromley (eds), British Social Attitudes: The 18th Report-Public Policy, Social Ties, London Sage.

Hills, J. (2002) “Following or Leading Public Opinion? Social Security Policy and Public Attitudes Since 1997” Fiscal Studies 23 539-58.

Leo T.W. (2005) “The Impact of Child Custody Legislation on Parental Labour Supply and Child Quality: A Differential Fecundity Perspective” mimeo University of Toronto Economics Department.

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 d'Economie Politique (Rouge et Cie, Paris).

Rao C.R. (1973) Linear Statistical Inference and Its Applications. Wiley

Rayner J.C.W and D.J.Best (1989) Smooth Tests of Goodness of Fit. Oxford University Press.
Reed, W. J., 2001, The Pareto, Zipf and other power laws, Economics Letters 74, 15-19.

Silverman B.W. (1986) Density Estimation for Statistics and Data Analysis. Monographs on Staatistics and Applied Probability. Chapman and Hall.

Smith A. (1776) An Enquiry Into the Nature and Causes of the Wealth of Nations. Liberty Classics

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.

Appendix 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) and Davidson and Duclos (1997), (2000)) are also all members of this class. These tests are often criticized for their potential inconsistency, that is of failing to reject a false null asymptotically, thus making the case for less powerful but non-the-less 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 T, the above class of tests can be represented by:

$$P = H(T)\sum_{k=1}^{K+1}\left(\sum_{j=1}^{k} I(j)\int_{z_{j-1}}^{z_j}(f(x)-g(x))dx\right)^2 G_k + O\left(\frac{1}{T}\right)$$

here z_i , i = 1,..,K together with $z_0 = a$ and $z_{K+1} = b$ represent ordered ($z_i < z_{i+1}$) points in the region [a,b] corresponding to a mutually exclusive and exhaustive partitioning of that region, G_k corresponds to 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, I(j) is an indicator function. The elements G_k are O(1) asymptotically and H(T), the sample size factor, is monotonic 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 test f(x) and g(x) are two empirical densities being compared, in both cases I(j) = 1 for j=k, = 0 otherwise (H(T) = T in the former and $T_fT_g/(T_f+T_g)$ in the latter where T_f and T_g are the sample sizes from the respective distributions). In 1st Order Stochastic Dominance (and Lorenz) tests f(x) and g(x) correspond to empirical density functions (or monotonic transformations of them) with 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 I(j) = 1.

Given the sample size factor is O(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:

$$(1) \qquad \int_{z_{k-1}}^{z_k} (f(x)-g(x))dx = 0 \quad for\ all\ z_k, \quad k = 1,..,K+1$$

The partition points z_i 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 z_i'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. For smooth, continuous functions f(x) and g(x) defined on [a,b] let there be M ordered interior intersection points w_i , such that f(x) = g(x), $x = w_i$, i = 1,..,M and $f(x) \neq g(x)$ otherwise, except possibly at x = a and x = b. Then the z_k satisfying (1) number at most K where K < M.

Proof. Suppose, without loss of generality, f(x) > g(x) for $x \in [a, w_1)$ then, from the smoothness and continuity assumptions for f(x) and g(x), |f(x)-g(x)| > 0 for $x \in (w_i , w_{i+1})$, i = 1,....,M-1. Since:

$$\int_a^{y(\leq w_1)} (f(x)-g(x))dx > 0$$

there can be no partition point in $[a, w_1)$, otherwise a term of $O(1)$ remains in the extreme left tail of the region [a,b]. Similarly since:

$$(-1)^I \int_{y(\geq w_M)}^{b} (f(x)-g(x))dx > 0$$

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

$$(-1)^i \int_{y_L}^{y_U} (f(x)-g(x))dx > 0 \ for \ all \ y_L < y_U, \ y_L, \ y_U \in (w_i, w_{i+1})$$

there can be at most one partition point in (w_i , w_{i+1}) for i = 1,...,M-1, otherwise a term of $O(1)$ remains within the region (w_i, w_{i+1}). Hence there are at most M-1 partition points satisfying (1).

Though the result is “simple” its practical implications are significant. For test inconsistency the set of points satisfying (1), 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 z_i’s, specifically not having a z_i at the poverty line, will engender inconsistency when comparisons are made over the whole distribution. Clearly the z_i'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.

Appendix 2. The Monte-Carlo Study.

Independent samples were drawn on f(x) ~ N(0,1) and g(x) ~ N(0, 2.25). The intersection points (ip) for these two distributions is given by ip $= \pm\sqrt{(\ln(2.25)*2.25/1.25)}$ {= 1.2082} and the exact value of the IA index for these two distributions is 0.1936. Bias to $O(h^2)$ in the kernel estimate f^e is given in Pagan and Ullah(1999) by:

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

Where h is the window width, μ_2 is the second moment of the kernel variate and for this case the second derivative of N(0, σ^2) is given by:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} \left(\frac{x^2}{\sigma^2} - 1\right)$$

In the region of the intersection in the example this, 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 IA since it always engenders positive bias in the estimate of the overlap measure (see below). 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. IA 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 equi-probable cells were conducted for both the intersection point estimates and IA statistic based upon known and estimated intersection points.

Table 1

n = 1000	Mean	Std Deviation	$\chi^2(8)$ Normality Test	P(Upper Tail)
Lower IP	-1.2412	0.1526	7.8	0.4532
Upper IP	1.2535	0.1522	3.8	0.8747
IA(IP known)	0.1916	0.0202	9.6	0.2942
IA(IP estimated)	0.1954	0.0189	8.7	0.3682
n = 5000				
Lower IP	-1.2271	0.0738	4.7	0.7891
Upper IP	1.2194	0.0709	4.7	0.7891
IA(IP known)	0.1941	0.009	5.2	0.736
IA(IP estimated)	0.195	0.0094	7.4	0.4942
n = 10000				
Lower IP	-1.2206	0.0569	6 .0	0.6472
Upper IP	1.2188	0.0604	4.3	0.8291
IA(IP known)	0.194	0.006	7.4	0.4942
IA(IP estimated)	0.1945	0.0058	17.8	0.0228

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 IA is not largely 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 its distributional properties for an arbitrarily defined set of points x_k. The index will be biassed downwards since the overlap measure 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 (see appendix). As indicated in the appendix, these dangers can be mitigated by choosing a larger number of points x_k than anticipated intersection points.

Two environments were investigated. The first, where distributions intersect once, is based upon two distributions N(a,1) and N(b,1) where d = b-a was varied from 0.1 to 1.5 in increments of 0.2. The second, where distributions intersect twice, 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 ranging from 500 to 2500 in increments of 500 and Alienation Indices calculated based upon k partitions with k set at 5, 10 and 20. Partition points were determined by equiprobable 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:

$$\ln\left(\frac{IA_i}{IA_i^{true}}\right) = \beta_0 + \beta_1\frac{1}{T_i} + \beta_2\frac{1}{K_i} + \beta_3\frac{1}{K_i^2} + \beta_4\frac{d_i}{T_i} + \beta_5\frac{1}{K_iT_i} + \beta_6\frac{d_i}{K_i} + \beta_7\ln(AI_i^{true}) + \beta_8(\ln(AI_i^{true}))^2 + \varepsilon_i$$

where T represents the sample size, K represents the number of partitions, AI corresponds to the estimated index, IA^{true} corresponds to the true value of the index and "i" corresponds to the i'th experiment. Since $IA = IA^{true} + \text{Bias}$ the equation can be seen as describing an approximation to the Bias / IA^{true} 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 the equation when IA is measured ignoring the intersection points and when IA is measured by incorporating the known intersection points in the partition structure. The three equation estimates are reported in Tables 2 and 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 $\beta_1/T + \beta_4 d/T + \beta_5/KT$. At the sample means these are -0.0816 and 0.00209 respectively for the E(p|p<q) 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 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)
$(1/\# \text{Partitions})^2$	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)
$\ln(IA^{true})$	-1.4408	0.1987	0.2097	-0.0110
		(0.0545)	(0.0587)	(0.0062)
$(\ln(IA^{true}))^2$	2.7483	0.0965	0.0982	-0.0016
		(0.0114)	(0.0123)	(0.0013)
R Squared		0.894	0.875	0.472
Sigma		0.061	0.065	0.007

Table 3. Variance Shift Effect.

Variable		Intersection points excluded	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)
Location Shift/Sample Size	0.0007	-145.1569	-136.9142	-8.2427
(1/# Partitions)(1/Sample		(19.9354)	(20.2351)	(2.1994)
Size)	0.0001	-363.9128	-331.4417	-32.4710
Scale Shift/# Partitions		(149.1667)	(151.4089)	(16.4571)
	0.0933	1.5124	1.2696	0.2428
ln(IA^{true})		(0.1841)	(0.1869)	(0.0203)
	-1.5357	0.1605	0.2152	-0.0547
(ln(IA^{true}))^2		(0.0785)	(0.0797)	(0.0087)
	2.8313	0.0880	0.0974	-0.0094
R Squared		(0.0168)	(0.0170)	(0.0018)
Sigma		0.864	0.845	0.801
		0.059	0.060	0.007

Asymptotic bias may be studied by allowing $T \to \infty$, in this case the increase in alienation brought about by increases in location and/or scale parameters increases the relative bias due to $E(p|p<q) \leq E(p)$ 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 $-\beta_2/K^2 - \beta_3/K^3 - \beta_6 d/K^2$, 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[7]. The latter phenomenon is not surprising, smaller intervals generally imply smaller approximation biasses 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^{true}) terms (the only ones that remain as 1/T and $1/K \to 0$) that

[7]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.

relative bias is present asymptotically. The absolute magnitude of the bias may be obtained by multiplying by IA^{true}, these biases are reported in Table 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 4. Approximate Absolute Asymptotic Bias (Large K)

IA^{true}	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