

# Polarization Measurement and Inference in Many Dimensions When Subgroups Can Not Be Identified

*Gordon Anderson*  
*University of Toronto*

**Abstract** The most popular general univariate polarization indexes for discrete and continuous variables are extended and combined to describe the extent of polarization between agents in a distribution defined over a collection of many discrete and continuous agent characteristics. A formula for the asymptotic variance of the index is also provided. The implementation of the index is illustrated with an application to Chinese urban household data drawn from six provinces in the years 1987 and 2001 (years spanning the growth and urbanization period subsequent to the economic reforms). The data relates to household adult equivalent log income, adult equivalent living space, which are both continuous variables and the education of the head of household which is a discrete variable. For this data set combining the characteristics changes the view of polarization that would be inferred from considering the indices individually.

Special Issue

[The Measurement of Inequality and Well-Being: New Perspectives](#)

**JEL** C14, C30, I32

**Keywords** Multivariate polarization measurement

**Correspondence** Gordon Anderson, Department of Economics, University of Toronto, Max Gluskin House, 150 St George St., Toronto, Ontario M5S 3G7, Canada, e-mail: [anderson@chass.utoronto.ca](mailto:anderson@chass.utoronto.ca)

**Citation** Gordon Anderson (2011). Polarization Measurement and Inference in Many Dimensions When Subgroups Can Not Be Identified. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 5, 2011-11. doi:10.5018/economics-ejournal.ja.2011-11. <http://dx.doi.org/10.5018/economics-ejournal.ja.2011-11>

## 1 Introduction

The functionings and capabilities approach to wellbeing measurement (Sen 1992) has given considerable impetus to multidimensional analyses of wellbeing (Grusky and Kanbur 2006). The argument is that individual wellbeing is not just a matter of the incomes they have or could achieve, among other things it depends on individual health and educational status, their political freedoms and environmental factors. In the absence of a well specified wellbeing aggregator of these many sensibilities (i.e. some form of utility function) evaluation of wellbeing has to be evaluated over these many dimensions which of course could be measured discretely or continuously.

The multivariate polarization measure presented here is founded upon the notion of polarization within a population distribution  $f(x)$  of individual characteristics  $x$  into potentially many possible groups which are not identified<sup>1</sup> a priori. Imagine for example a population which is a mixture of  $K$  classes with respective distributions  $f_k(x)$  and proportions  $w_k$  so that the population distribution  $f(x)$  may be written as:

$$f(x) = \sum_{k=1}^K w_k f_k(x)$$

When no class identifier or information on agent membership of the  $f_k(x)$ 's (i.e. the sub group distributions) is available, all that is observed is  $f(x)$ , the population distribution. This is the unidentified case for which the polarization measures discussed herein are appropriate. Sometimes, given additional information, the sub distributions can be estimated facilitating calculation of the probability (or partial identification) of group membership for an individual with characteristics  $x$  or indeed perfect stratification where group membership is known with probability 1. For example Anderson, Pittau and Zelli (2011) posit that the incomes of each class are driven by distinct stochastic processes which precipitate distinct log normal distribution specifications for the  $f_k(x)$ 's permitting, through the employment of semi parametric techniques, estimation of all of the parameters of the mixture distribution. This permits estimation of the extent of polarization

---

<sup>1</sup> Here "identification" refers to known membership of a group rather than a sense of kinship or proximity to other members in a group which is the sense in which it will be used later in defining the polarization measure.

between any two classes in a similar fashion to when class membership is completely identified. Here no such information is available.

Esteban and Ray (1994) and Duclos, Esteban, and Ray (2004) posited a collection of propositions with which such a Polarization measure for the unidentified case should be consistent and proposed a collection of univariate measures appropriate for a variety of circumstances that would reflect such polarization between potentially many groups. The propositions are based upon a so-called Cohesion (or Identification) and Alienation nexus wherein notions of polarization are fostered jointly by an agent's sense of increasing within-group identity or association and between-group distance or alienation.

There have been several proposed univariate polarization indices which focus on an arbitrary number of groups<sup>2</sup> in this unidentified case (Esteban and Ray, 1994; Esteban, Gradin and Ray, 1998; Zhang and Kanbur, 2001; Duclos, Esteban and Ray 2004) and a similar number that focus on just two identified groups i.e. when the sub distributions above are observed (Alesina and Spolaore 1997; Foster and Wolfson 1992; Wolfson 1994; Wang and Tsui, 2000). Anderson, 2004 considers tests for various types of polarization between two identified or partially identified groups based upon the anatomy of their respective distributions or the mixture of their respective distributions when the groups were only partially identified. While much work has been done on extending one dimensional wellbeing measures to many dimensions in the context of poverty (Duclos, Sahn and Younger 2006) and inequality measurement (Maassoumi 1986, 1999, Koshevoy and Mosler 1997, Tsui 1995 and Anderson 2008)<sup>3</sup> little has been done in extending polarization measures to the many dimensioned case. While Gigliarano and Mosler, (2009) develop a family of multivariate polarization measures based upon measures of between and within group multivariate variation and relative group size which exploit notions of subgroup decomposability and Anderson (2010) and Anderson, Linton and Leo (2011) have developed a trapezoidal measure of polarization which can be applied to two identifiable groups or within a population distribution provided at least two modal points are identified (i.e. the partially identified case), multivariate polarization measures have not been developed for the more general non-identified many group case, nor

---

<sup>2</sup> And a fortiori two groups.

<sup>3</sup> All however confine themselves to continuous variables.

for the case where the joint distribution of sensibility indicators is a mixture of discrete and continuous variables.<sup>4</sup>

An excellent summary of the properties of the univariate indices is to be found in (Esteban and Ray, 2007) wherein the properties of indices are evaluated in terms of their coherence with some basic axioms that reflect three broad notions, 1) When there is only one group there is little polarization, 2) polarization increases when within group inequality is reduced, 3) polarization increases when between group inequality increases. The axioms are formed around a notional univariate density that is a mixture of kernels  $f(x, a)$  that are symmetric uni-modal on a compact support of  $[a, a+2]$  with  $E(x) = \mu = (a+1)$  also representing the mean or mode. However these axioms are readily extended to multivariate densities of continuous variables by thinking in terms of a notional multivariate density that is a mixture of multivariate kernels so that  $x$  is simply a  $j$  dimensional vector. The kernels are subject to slides (location shifts)  $g(y) = f(y-x)$ , which may be contemplated in terms of the Euclidean distance<sup>5</sup> between vectors  $y$  and  $x$ , and squeezes (shrinkages) of the form  $f^\lambda(x) = f(\{x - [1-\lambda]\mu\} / \lambda) / \lambda$  ( $0 < \lambda < 1$ ) where now  $\mu$  is a  $j$  dimensional vector of means or modal values of the multidimensional kernel  $f(x, a)$  that is symmetric on a compact support of  $[a, a+2]$  where  $a$  is a  $j$  dimensional vector.

Potential indices are evaluated in the context of such changes in terms of the extent to which they satisfy a set of axioms which reflect the following set of ideas. The squeeze of a uni-modal distribution cannot increase polarization and symmetric squeezes of the two kernels cannot reduce polarization. Sliding two kernels away from one another increases polarization and common population scaling preserves the polarization ordering. Polarization indices have to come from a family where if  $x$  and  $y$  are independently distributed with marginal distributions  $f(x)$  and  $f(y)$  then the index is the expected value of some function  $T(f(x), |x-y|)$  which is increasing in its second argument. Symmetric squeezes of the sub distributions weakly increases polarization. The index should be non-monotonic with respect to outward slides of the sub distributions and flipping the distribution around its support should leave polarization unchanged. Most of these ideas can be

---

<sup>4</sup> Furthermore extensions of the stochastic dominance techniques introduced in Anderson (2004), which really explore the anatomy of polarizing distributions, would prove cumbersome in many dimensions because it is not obvious how to define a sensible partition of the distribution across those many dimensions.

<sup>5</sup> Other distance metrics (for example Mahalanobis 1936 or Bregman 1967) could be employed.

contemplated with respect to multivariate densities of continuous variables though there is some difficulty when multivariate densities of discrete variables are contemplated unless slides of the discrete outcome values are permitted and squeezes of the distributions contemplated in terms of transfers of mass between outcome values.

Here the most popular general univariate polarization indices for discrete (Esteban and Ray 1994), and continuous (Duclos, Esteban and Ray 2004) variables are combined and extended to describe the extent of polarization between agents in a distribution defined over a collection of many discrete and continuous agent characteristics. The univariate indices have been demonstrated to satisfy the aforementioned axioms. The implementation of the index is illustrated with an application to Chinese urban household data drawn from six provinces in the years 1987 and 2001 (years spanning the growth and urbanization period subsequent to the economic reforms). The data relates to household adult equivalent log income, adult equivalent living space, which are both continuous variables and the education of the head of household which is a discrete variable.

## 2 The Extension to Many Variables both Discrete and Continuous

The multivariate generalization of the Duclos, Esteban, and Ray (2004) (DER) Polarization index is, like DER, based upon the sample equivalents of the population concepts. For scalar continuous  $x$  with distribution function  $F(x)$  the DER index is given by:

$$P_{\alpha} = \int \int f(x)^{\alpha} |y - x| dF(y) dF(x) \quad [1]$$

Some intuition for the index may be gained by thinking in terms of  $f(x)^{\alpha}$  as the degree of identification or cohesion experienced by an agent with income  $x$  ( $f(x)$  being an indicator of mass around  $x$ ) and  $a(x)$  as the degree of alienation experienced by a person with income  $x$  where:

$$a(x) = \int |y - x| dF(y)$$

$p_a(x) = f(x)^\alpha a(x)$ , the area of a rectangle with height  $f(x)^\alpha$  and base  $a(x)$ , is then the degree of polarization experienced by an agent with  $x$  and [1] corresponds to the average polarization experienced across the population of agents.

DER, (in Duclos, Esteban and Ray 2004a) demonstrate the estimator of [1] to be asymptotically normally distributed with an asymptotic variance  $V$  given by:

$$V_\alpha = \text{var}_{f(y)} \left( (1+\alpha)f(y)^\alpha \int_0^\infty \|y-x\| dF(x) + y \int_0^\infty f(x)^\alpha dF(x) + 2 \int_y^\infty \|y-x\| f(x)^\alpha dF(x) \right) \quad [2]$$

Their development of the variance formula is sketched in the Appendix. A similar discrete variable index is provided in Esteban and Ray (1994) and is given by:

$$P_\alpha = K \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \pi_i^{1+\alpha} \pi_j$$

where  $\pi_i$  is the sample weight of the  $i$ 'th observation and  $K$  is a normalizing factor.

Development of the polarization index was founded on a set of axioms that such an index should obey, the axioms concern changes (squeezes and slides) in the uni-modal sub distributions in the mixture distribution that is  $f(x)$ . The resultant index reflects the two primary factors that underlay polarization, the alienation or distance between groups (given by  $|y-x|$ ) and the association within a group (given by  $f(x)^\alpha$ ). Indeed the intuitive interpretation of  $P_\alpha$  as the average value of the areas of all possible trapezoids that can be formed under  $f(x)$  whose average height is  $f(x)^\alpha$  and whose base is  $|x-y|$  can be related to the trapezoidal index of polarization employed in Anderson (2010) to study multivariate poverty states and in Anderson, Leo and Linton (2011) to study multivariate convergence issues. Here  $\alpha$  is a polarization sensitivity parameter<sup>6</sup> chosen by the investigator such that  $0.25 \leq \alpha \leq 1$  with higher values of  $\alpha$  corresponding to increased sensitivity. The same axioms can be applied when  $x$  is a vector and where  $\|x-y\|$  is the Euclidean distance between the vectors.<sup>7</sup>

<sup>6</sup> Note when  $\alpha = 0$  the index is in essence twice the Gini coefficient thus a similar value in the following would provide a multivariate version of a Gini like coefficient and its variance.

<sup>7</sup> Anderson, Crawford and Leicester (2011) employ Euclidian distance in developing a non-parametric approach to multivariate welfare rankings.

Let  $w_i$  and  $z_i$  be jointly distributed vectors describing the status of the  $i$ 'th agent with  $w_i$  being a  $k \times 1$  vector of continuous variables and  $z_i$  being an  $h \times 1$  vector of continuous variables with  $i=1, \dots, n$  being the elements of the sample. The continuous variables all reflect wellbeing positively and for convenience are defined on  $R_k^+$  and the discrete variables are ordered integers reflecting positive wellbeing in the same fashion.<sup>8</sup> The joint density of the  $w$ 's for a given configuration of  $z$ 's is  $f_z(w|z)$  and the joint probability of the  $z$ 's is  $p(z)$  so that the joint density of the  $w$ 's and  $z$ 's for the  $i$ 'th agent with continuous characteristics  $w_i$  and discrete characteristics  $z_i$  is given by  $f(w_i, z_i) = f_i(w_i | z_i) p(z_i)$  which corresponds to her degree of identification. As for the alienation component let  $x_i$  be the stacked vector  $w_i | z_i$  then the dimension normalized Euclidean distance<sup>9</sup> between agents  $i$  and  $j$  given by  $\|x_i - x_j\|$  is well defined and may be written as:

$$\|x_i - x_j\| = \sqrt{\frac{\sum_{q=1}^Q (x_{iq} - x_{jq})^2}{Q}}$$

where  $x_{iq}$  is the  $q$ 'th element of the vector  $x_i$  where  $Q = k+h$ . For notational convenience denote the first  $k$  continuous components of the vector  $x$  as  $x_{\{c\}}$ .

Then, retaining the trapezoidal intuition, a multivariate version of [1] is given by:

$$P_\alpha = \sum_{z \in X} \int_{R_k^{w_x}} \sum_{z \in Y} \int_{R_h^{w_y}} (f_z(w | z) p(z))^\alpha \|y - x\| dF(w_y) dF(w_x) \quad [1a]$$

Here summation is over the domain of each element of the  $z$  vector and integration is over the domain of each element of the  $w$  vector. As in the univariate case the alienation or distance between groups is given by  $\|y-x\|$  and the association within

---

<sup>8</sup> For example the continuously measured variables may represent levels of consumption, leisure and housing stock whereas the discretely measured variables may reflect levels of educational, health or freedom status.

<sup>9</sup> In DER the columns of  $X$  are mean standardized and assumed to reside in the positive orthant which has been followed here, however other distance measures could be equally well employed, for example Mahalanobis distance (Mahalanobis 1936) or Bregman distance (Bregman 1967) would probably better accommodate the variations over the different dimensions (I'm grateful to an anonymous referee for pointing this out).

a group given by  $f(w,z)^\alpha$  in exactly the same fashion.<sup>10</sup> By employing kernel estimates of the conditional multivariate distributions and sample estimates of the population proportions  $p(z)$  the sample equivalents, given  $n$  observations on  $Q$  variables in an  $n \times Q$  matrix  $X$  with typical element  $x_{iq}$   $i = 1, \dots, n$ ,  $q = 1, \dots, Q$  and typical row  $x_i$  the index can be seen to be:

$$\widehat{P}_\alpha = \frac{\sum_{i=1}^n \widehat{(f(w_i | z_i) p(z_i))^\alpha} \sum_{j=1}^n \sqrt{\sum_{q=1}^Q (x_{iq} - x_{jq})^2}}{n^2 \sqrt{Q}}$$

The multivariate version of [2], the variance of index is given by:

$$V_\alpha = \text{var}_{f(y)} \left( \begin{array}{l} (1 + \alpha)(f(w | z)P(z))^\alpha \sum_z \int_0^\infty \|y - x\| dF(w | z)P(z) + \\ \|y\| \sum_z \int_0^\infty (f(w | z)P(z))^\alpha dF(w | z)P(z) \\ + 2 \sum_z \int_y^\infty \|y - x\| (f(w | z)P(z))^\alpha dF(w | z)P(z) \end{array} \right) \quad [2a]$$

Where after ordering the vectors  $x_i$  on  $\|x_i\|$  as  $x_i^0$ , the first, second and third terms of the  $i$ 'th element of the variance vector may be respectively estimated in an obvious fashion as:

---

<sup>10</sup> Note that Esteban and Ray (1994) and DER respectively offer different ranges for  $\alpha$  for discrete univariate and continuous univariate distributions this can be accommodated in the present context by considering the association component as  $f(w|z)^{\alpha c} p(z)^{\alpha d}$  where  $\alpha c$  is the polarization parameter for the continuous components and  $\alpha d$  is the polarization parameter associated with the discrete components.



$$\frac{\overbrace{f(w_i | z_i)P(z_i)}^\alpha \sum_{j=1}^n \sqrt{\sum_{k=1}^Q (x_{ik} - x_{jk})^2}}{n\sqrt{Q}}$$

$$\sqrt{\frac{\sum_{k=1}^Q x_{ik}^2}{Q} \frac{\sum_{j=1}^n \overbrace{f(w_i | z_i)P(z_i)}^\alpha}{n}}$$

$$2 \frac{\overbrace{f(w_i | z_i)P(z_i)}^\alpha \sum_{j=i}^n \sqrt{\sum_{k=1}^Q (x_{ik}^o - x_{jk}^o)^2}}{n\sqrt{Q}}$$

Essentially the generalization simply involves employing the dimension normalized Euclidean norm for  $|y-x|$  and  $|y|$  when they are  $Q$  dimensional vectors together with multivariate kernel estimates of  $f(w|z)p(z)$  for  $f(x)$  and  $f(y)$  raised to an appropriate power value of  $\alpha$ , the polarization sensitivity index, which is of course the choice of the investigator.

### 3 An Application: Chinese Urban Households 1987-2001

There is a suspicion that the economic reforms and one child policy initiated in China in the late 1970's together with the massive urbanization over the period changed the nature of urban households and families. The One Child Policy (OCP) intervention changed fundamentally the nature of both existing and anticipated marriage arrangements and influenced family formation decisions in many dimensions. Anderson and Leo (2007), in studying the impact of the policy on family formation in urban China construed the OCP as a rationing policy constraining the quantity (but not the quality) of children, evidence of increased positive assortative pairing of couples was observed as was increased investment in children and, also consistent with rationing theory, income became less of a factor in determining family size though it did become an increasingly important

determining factor in investment in children. At the same time there was an unprecedented growth in household incomes largely attributed to the economic reforms (the average annual growth rate of city incomes over the period 1990-1999 was over 18%, Anderson and Ge 2004) and a massive migration to the cities (in 1985 20% of the Chinese population was urbanized, by 1999 over 42.6% was urbanized, Anderson and Ge 2005). All of which could have changed substantially the way that households relate to one another, one aspect of which is the extent to which households are polarized.

Table 1 presents the summary statistics of data on two independent surveys of urban households from three coastal and three interior provinces<sup>11</sup> in China for the years 1987 (for which there were 3651 observations) and 2001 (for which there were 4297 observations) a period over which the reforms took effect. The data were used to generate observations on log adult equivalent household income (at constant prices), adult equivalent<sup>12</sup> living space (in square meters) and an integer index of the education level of the head of household. Thus in this example the household is the agent.

The marginal distributions of the three polarization characteristics exhibit quite distinct structures and changes over the period. When these variables are put together in a joint distribution its anatomy is likely to change over the period and given its central role in the polarization calculus the nature of polarization is likely to change. A considerable reduction in family size (the effect of the one child policy) and considerable increases in both equivalent incomes and living space (due in part to growth and in part to reductions in family size) and educational attainment are evident. The variation in incomes, living space and education also increased over the period (suggesting a diverging society) whereas the family size was clearly converging. Incomes and education are negatively skewed (long upper tail) incomes increasingly so and education decreasingly so over the period. Living space is positively skewed and decreasingly so over the period. Family size actually switched skew from negative to positive over the period. With the exception of family size and education, and family size and income, correlations

---

<sup>11</sup> The coastal provinces were Jilin, Shandong and Guangdong the interior, Sichuan, Shaanxi and Hubei .

<sup>12</sup> Equalization was effected using the square root rule (Brady and Barber 1948).

Table 1. Summary Statistics (1987 n=3651, 2001 n = 4297)

	1987	2001	1987	2001	1987	2001	1987	2001
	Equiv log income	Equiv log income 1987 prices	Equiv living space (square meters)	Equiv living space (square meters)	Edu of House Head	Edu of House Head	Family Size	Family Size
Mean	4.8227	8.8212	17.1178	24.7200	3.1035	3.4263	3.6935	3.0856
Median	4.8579	8.9074	15.2053	22.5167	4.0000	4.0000	4.0000	3.0000
StdDev	0.4194	0.8489	9.4884	12.6218	1.5754	1.6137	0.9798	0.7608
Skew	-0.2518	-0.3046	0.6047	0.5237	-1.7072	-1.0666	-0.9385	0.3375

1987 correlations

	log income	living space	education
Living space	0.048729790		
Education	-0.054868589	-0.043417229	
Family size	-0.030030869	-0.035464050	0.10348972

2001 correlations

	log income	living space	education
Living space	0.19662306		
Education	-0.18139926	-0.13138635	
Family size	0.019805711	-0.067196345	0.080520251

between the variables have become uniformly stronger over the period. The family size/income correlation has actually switched sign as though children have changed from being an inferior good to being a normal or luxury good over the period.

To calculate the polarization statistic the continuous multivariate mean standardized pdf's were estimated using a multivariate standard normal kernel with a window width  $h = 1.06 \cdot \sigma(x) \cdot n^{-1/(4+k)}$  (Silverman 1986). The seven outcome educational scale was condensed to a three outcome scale, 1, 2 and 3 corresponding to high, medium and low educational attainments. Table 2 reports the univariate polarization indices and standard errors for the continuous and discrete measures as per DER, Table 3 reports the paired multivariate measures and Table 4 reports the overall multivariate measures.

For all values of the polarization sensitivity parameter the index shows increases for all income, house space and education variables and, based upon the samples in the two years being independent of one another, the increase is seldom insignificant at usual levels of significance though the differences do appear to diminish as the polarization sensitivity parameter increases. Increasing the sensitivity parameter can be interpreted as increasing the relative influence of the identification component relative to the inequality or alienation component suggesting that polarization changes in the univariate case are largely a result of changes in the alienation component.

The joint pair-wise distributions reported in Table 3 exhibit quite different effects to the univariate cases. At low levels of polarization sensitivity significant polarization is still the norm for all pair-wise comparisons with depolarization being the norm in almost all cases and significantly so as higher orders of polarization sensitivity are considered. Thus pairing or combining variables appears to dilute the alienation effect even further.

Turning to the polarization measures across all three characteristics which, together with a test for depolarization, are reported in Table 4, note that the null of depolarization is never rejected for all levels of polarization sensitivity.

Furthermore the differences, which are now reductions in polarization, are more substantial the more heavily weighted is the identification component. So it appears that expanding the dimensions over which polarization is considered further denudes the polarization observed when the various household characteristics are considered individually.

*Table 2. Univariate Measures (Standard Errors in Brackets)*

Sensitivity Parameter ( $\alpha$ )	Income 1987	Income 2001	Housing 1987	Housing 2001	Edu 1987	Edu 2001
0.25	0.1173 (0.0003)	0.1321 (0.0003)	0.4182 (0.0007)	0.4314 (0.0007)	0.1784 (0.0016)	0.2166 (0.0021)
0.5	0.1524 (0.0003)	0.1661 (0.0003)	0.3634 (0.0004)	0.3745 (0.0004)	0.1339 (0.0011)	0.1641 (0.0015)
0.75	0.2041 (0.0003)	0.2136 (0.0003)	0.3271 (0.0003)	0.3349 (0.0002)	0.1019 (0.0007)	0.1244 (0.0012)
1.0	0.2789 (0.0004)	0.2793 (0.0003)	0.3007 (0.0002)	0.3051 (0.0002)	0.0785 (0.0005)	0.0942 (0.0009)

*Table 2a.* Univariate Polarization Tests [ $H_0: Pol_{1987}-Pol_{2001} \geq 0$ , “t”, (P(t < “t”))]

Sensitivity Parameter ( $\alpha$ )	Income	Housing	Education
0.25	-34.5502 (0.0)	-13.4451 (0.0)	-14.0846 (0.0)
0.5	-32.8214 (0.0)	-20.8722 (0.0)	-15.6281 (0.0)
0.75	-20.5224 (0.0)	-21.5707 (0.0)	-16.0812 (0.0)
1.0	-0.7099 (0.2389)	-15.9004 (0.0)	-15.3907 (0.0)

*Table 3* Multivariate Polarization: Paired Comparisons

Sensitivity Parameter ( $\alpha$ )	Income and Housing 1987	Income and Housing 2001	Income and Edu 1987	Income and Edu 2001	Housing and Edu 1987	Housing and Edu 2001
0.25	0.4136 (0.0011)	0.4170 (0.0007)	0.2984 (0.0007)	0.3325 (0.0009)	0.3578 (0.0008)	0.3697 (0.0007)
0.5	0.4900 (0.0011)	0.4735 (0.0007)	0.3146 (0.0007)	0.3310 (0.0008)	0.2574 (0.0006)	0.2545 (0.0005)
0.75	0.6122 (0.0011)	0.5600 (0.0008)	0.3419 (0.0006)	0.3340 (0.0008)	0.1904 (0.0005)	0.1778 (0.0003)
1.0	0.7913 (0.0012)	0.6810 (0.0009)	0.3806 (0.0006)	0.3407 (0.0008)	0.1439 (0.0003)	0.1255 (0.0002)

*Table 3a.* Bivariate Polarization Tests [ $H_0: Pol_{1987}-Pol_{2001} \geq 0$ , “t”, (P(t < “t”))]

Sensitivity Parameter ( $\alpha$ )	Income and Housing	Income and Education	Housing and Education
0.25	-2.5810 (0.0049)	-29.4864 (0.0000)	-10.9553 (0.0000)
0.5	12.4364 (1.0000)	-15.2877 (0.0000)	3.7932 (0.9999)
0.75	37.2647 (1.0000)	7.5415 (1.0000)	21.7465 (1.0000)
1.0	73.1129 (1.0000)	38.5854 (1.0000)	41.5799 (1.0000)

Table 4. Multivariate Measures (Standard Errors in Brackets)

Sensitivity Parameter $\alpha$	Income, Housing and Edu 1987	Income, Housing and Edu 2001	Polarization test [ $H_0: Pol_{1987} - Pol_{2001} \geq 0$ , “t”, (P(t < “t”))]
0.25	0.3993 (0.0010)	0.3994 (0.0009)	-0.0616 (0.4754)
0.5	0.3932 (0.0010)	0.3602 (0.0009)	25.1768 (1.0000)
0.75	0.4043 (0.0010)	0.3332 (0.0009)	53.2727 (1.0000)
1.0	0.4295 (0.0010)	0.3141 (0.0009)	83.9697 (1.0000)

## 4 Conclusions

Many researchers have argued that, in the absence of a plausible aggregator of the many factors that affect wellbeing, its measurement needs to be pursued in the context of the several variables available rather than relying on just one of them. This applies to most aspects of wellbeing measurement. Here, by combining multivariate versions of the Polarization indices developed in Esteban and Ray (1994) and Duclos, Esteban, and Ray (2004) the polarization measurement toolkit has been extended to the case where the status of an agent is represented by many characteristics which can be both discretely and continuously measured and the agent subgroups in a population are not identified. The asymptotic variance of the statistic has been provided to facilitate inference.

As an example the statistic was applied to Data on two independent surveys of urban households from three coastal and three interior provinces in China for the years 1987 and 2001, a period over which economic and family size reforms took effect and a period over which there was extensive urbanization. The data reflected the log adult equivalent household income (at constant prices), adult equivalent living space (in square meters) and an integer index of the education level of the head of household each of which may be construed as contribution to the wellbeing of the household.

The results, while obviously specific to these particular data, were salutary with regard to the use of univariate as opposed to multivariate polarization indices. While the individual univariate indices all reflected significant increases in polarization between households over the period of the reforms, when they were combined the polarization result was attenuated. For pair-wise combinations of the variables significant polarization was detected at low levels of polarization sensitivity but at high levels of polarization sensitivity significant depolarization was detected. When all three variables were combined in an index, significant depolarization was detected at all levels of polarization sensitivity.

### **Acknowledgments**

Many thanks Oliver Linton for help with multivariate Kernel Estimation and to Jean-Yves Duclos, Joan Esteban and two anonymous referees for helpful comments on earlier drafts of this paper.

## References

- Alesina, A., and E. Spolaore (1997). On the Number and Size of Nations. *Quarterly Journal of Economics* 112: 1027–1056.  
<http://ideas.repec.org/a/tpr/qjecon/v112y1997i4p1027-56.html>
- Anderson, G.J. (2004). Toward an Empirical Analysis of Polarization. *Journal of Econometrics* 122: 1–26.  
<http://ideas.repec.org/a/eee/econom/v122y2004i1p1-26.html>
- Anderson, G.J. (2008). The Empirical Assessment of Multidimensional Welfare, Inequality and Poverty: Sample Weighted Generalizations of the Kolmogorov-Smirnov Two Sample Test for Stochastic Dominance. *Journal of Economic Inequality* 6: 73–87.
- Anderson, G.J. (2010). Polarization of the Poor: Multivariate Relative Poverty Measurement Sans Frontiers. *Review of Income and Wealth* 56: 84–101.  
[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1551556](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1551556)
- Anderson, G.J., and Y. Ge (2004). Do Economic Reforms Accelerate Urban Growth: The Case of China. *Urban Studies* 11: 2197–2210.  
<http://usj.sagepub.com/content/41/11/2197.abstract>
- Anderson, G.J., and Y. Ge (2005). The Size Distribution of Chinese Cities. *Regional Science and Urban Economics* 35: 756–776.  
<http://ideas.repec.org/a/eee/regeco/v35y2005i6p756-776.html>
- Anderson, G.J., I. Crawford, and A. Leicester (2011). Welfare Rankings from Multivariate Data, a Nonparametric Approach. *Journal of Public Economics* 95: 247–252. <http://ideas.repec.org/a/eee/pubeco/v95y2011i3-4p247-252.html>
- Anderson, G.J., and T.W. Leo (2007). Family Formation and the One Child Policy in Urban China. Mimeo, Department of Economics, University of Toronto.
- Anderson, G.J., O. Linton, and T.W. Leo (2011). Rich Country-Poor Country Convergence from a Polarization-Cohesion Perspective. Mimeo, Department of Economics, University of Toronto.
- Anderson, G.J., G. Pittau, and R. Zelli (2011). Partially Identified Poverty Status: A New Approach to Measuring Poverty and the Progress of the Poor. Working Paper tecipa-421, Department of Economics, University of Toronto.  
<http://ideas.repec.org/p/tor/tecipa/tecipa-421.html>



- Brady D.S., and H.A. Barber (1948). The Pattern of Food Expenditures. *Review of Economics and Statistics* 30: 198–206. <http://www.jstor.org/stable/1926749>
- Bregman, L.M. (1967). The Relaxation Method of Finding the Common Points of Convex Sets and its Application to the Solution of Problems in Convex Programming. *USSR Computational Mathematics and Mathematical Physics* 7: 200–217.
- Duclos, J.-Y., J. Esteban, and D. Ray (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica* 72: 1737–1772. <http://ideas.repec.org/a/ecm/emetrp/v72y2004i6p1737-1772.html>
- Duclos, J.-Y., J. Esteban, and D. Ray (2004a). Polarization: Concepts, Measurement, Estimation. Bureau for Research in Economic Analysis of Development, Working Paper No. 059, April.
- Duclos, J.-Y., D. Sahn, and S.D. Younger (2006). Robust Multidimensional Poverty Comparisons. *The Economic Journal* 116: 943–968. <http://ideas.repec.org/a/ecj/econjl/v116y2006i514p943-968.html>
- Esteban, J.-M., and D. Ray. (1994). On the Measurement of Polarization. *Econometrica* 62: 819–851. <http://ideas.repec.org/a/ecm/emetrp/v62y1994i4p819-51.html>
- Esteban, J., and D. Ray (2007). A Comparison of Polarization Measures. UFAE and IAE Working Papers 700.07, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC). <http://ideas.repec.org/p/aub/autbar/700.07.html>
- Esteban, J., C. Gradin, and D. Ray (1998). Extensions of a Measure of Polarization, with an Application to the Income Distribution of Five OECD Countries. Mimeo, Instituto de Analisis Económico.
- Foster, J., and M. Wolfson (1992). Polarization and the Decline of the Middle Class in Canada and the USA. Mimeo, Vanderbilt University.
- Gigliarano, C., and K. Mosler (2009). Constructing Indices of Multivariate Polarization. *Journal of Economic Inequality* 7: 435–460. <http://ideas.repec.org/a/kap/jecinq/v7y2009i4p435-460.html>
- Grusky, D., and R. Kanbur (eds.) (2006). *Poverty and Inequality*. Stanford, CA: Stanford University Press.

- Koshevoy, G.A., and K. Mosler (1997). Multivariate Gini Indices. *Journal of Multivariate Analysis* 60: 252–276.  
<http://ideas.repec.org/a/eee/jmvana/v60y1997i2p252-276.html>
- Maasoumi, E. (1986). The Measurement and Decomposition of Multidimensional Inequality. *Econometrica* 54: 771–779.  
<http://ideas.repec.org/a/ecm/emetrp/v54y1986i4p991-97.html>
- Maasoumi, E. (1999). Multidimensioned Approaches to Welfare Analysis, chapter 15. In J. Silber (ed.), *Handbook of Income Inequality Measurement*. Norwell, MA: Kluwer Academic Publishers.
- Mahalanobis, P.C. (1936). On the Generalised Distance in Statistics. *Proceedings of the National Institute of Sciences of India* 2 (1): 49–55.
- Sen, A.K. (1992). *Inequality Reexamined*. Cambridge, MA: Harvard University Press.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman Hall.
- Tsui, K.-Y. (1995). Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson-Kolm-Sen Approach. *Journal of Economic Theory* 67: 251–265.  
<http://www.sciencedirect.com/science/article/pii/S0022053185710733>
- Wang, Y. Q., and K. Y. Tsui (2000). Polarization Orderings and New Classes of Polarization Indices. *Journal of Public Economic Theory* 2: 349–363.  
<http://onlinelibrary.wiley.com/doi/10.1111/1097-3923.00042/pdf>
- Wolfson, M.C. (1994). When Inequalities Diverge. *American Economic Review* 84: 353–358.  
<http://ideas.repec.org/a/aea/aecrev/v84y1994i2p353-58.html>
- Zhang, X., and R. Kanbur (2001). What Difference Do Polarization Measures Make? An Application to China. *Journal of Development Studies* 37: 85–98.  
<http://ideas.repec.org/a/taf/jdevst/v37y2001i3p85-98.html>

## Appendix

The derivation of the variance of the estimator follows DER (2004a) where  $P$  is rewritten in terms of  $a(y)$  and  $p_a(y)$  where:

$$a(y) = \int |y - x| dF(x) \text{ and } p_a(y) = f(y)^\alpha a(y)$$

Noting that  $a(y)$  may be written in terms of the population mean  $\mu$  and first partial moment  $\mu^*$  as:

$$a(y) \equiv \mu + y(2F(y) - 1) - 2\mu^*(y)$$

$$\text{where } \mu = \int y dF(y) \text{ and } \mu^*(y) = \int_{-\infty}^y x dF(x)$$

DER(2004a) show that, for large  $n$ ,  $P$  may be decomposed across its sources of sampling variability into:

$$\begin{aligned} P_\alpha(\widehat{F}) - P_\alpha(F) &\approx \int (\widehat{f}(y)^\alpha - f(y)^\alpha) a(y) dF(y) \\ &\quad + \int f(y)^\alpha (\widehat{a}(y) - a(y)) dF(y) \quad [A1] \\ &\quad + \int p_\alpha(y) d(\widehat{F} - F)(y) \end{aligned}$$

Assuming that the second order moments of all the components of the right hand terms are finite and that  $h$ , the window width of the kernel estimator, goes to 0 as  $n$  gets large, each term on the right hand side can be shown to be  $o(n^{-0.5})$  and each respectively may be written as:

$$\begin{aligned} & \frac{\alpha}{n} \sum_{i=1}^n (p_a(y_i) - P_\alpha); \\ & \frac{1}{n} \sum_{i=1}^n \left( \int f(y)^\alpha dF(y) y_i - 2y_i \int_{y_i}^{\infty} f(y)^\alpha dF(y) + 2 \int_{y_i}^{\infty} y f(y)^\alpha dF(y) \right) - \int f(y)^\alpha (\mu + 2yF(y) - 2\mu^*(y)) dF(y); \\ & \frac{1}{n} \sum_{i=1}^n (f(y_i)^\alpha a(y_i)) - P_\alpha \end{aligned} \quad [A2]$$

Applying the law of large numbers to [A1] and noting that root n times [A1] has expectation 0. Invoking the central limit theorem and collecting and re-arranging the terms yields the variance formula [2].

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

<http://dx.doi.org/10.5018/economics-ejournal.ja.2011-11>

The Editor