University of Toronto
Department of Economics

Working Paper 521

A new approach to measuring and studying the characteristics of class membership: The progress of poverty, inequality and polarization of income classes in urban China.

By Gordon Anderson, Alessio Farcomeni, Grazia Pittau and Roberto Zelli

October 20, 2014

A new approach to measuring and studying the characteristics of class membership: The progress of poverty, inequality and polarization of income classes in urban China.

Gordon Anderson[*], Alessio Farcomeni[†],
Maria Grazia Pittau and Roberto Zelli[‡]

March 19, 2014

Abstract

Classifying agents into subgroups in order to measure the plight of the "poor", "middle class" or "rich" is common place in economics, unfortunately the definition of class boundaries is contentious and beset with problems. Here a technique based on mixture models is proposed for surmounting these problems by determining the number of classes in a population and estimating the probability that an agent belongs to a particular class. All of the familiar statistics for describing the classes remain available and the possibility of studying the determinants of class membership is raised. As a substantive illustration we analyze household income in Urban China in the last decade of the 20th Century. Four income groups are classified and the progress of those "poor", "lower middle", "upper middle" and "rich" classes are related to household and regional characteristics to study the impact of urbanization and the one child policy on class membership over the period.

Keywords: Poverty Frontiers, Mixture Models, Class membership, Urban China.

JEL classification: C14; I32; O1.

[*]Department of Economics, University of Toronto, email: anderson@chass.utoronto.ca

[†]Department of Public Health and Infectious Diseases, Sapienza University of Rome, email: alessio.farcomeni@uniroma1.it

[‡]Department of Statistical Sciences, Sapienza University of Rome, email: grazia.pittau@uniroma1.it , roberto.zelli@uniroma1.it .

1 Introduction.

A cursory survey of leading economic journals will reveal a long established practice in the economics profession of classifying agents within a society into groups in order to measure and study their wellbeing or behavior. Invariably this has involved specifying boundaries or frontiers for class inclusion and exclusion purposes, (to establish who constitute the poor, the middle and the rich classes for example). To the extent that these boundaries have an arbitrary quality they have been a matter of much concern and dispute among researchers. Here a technique is proposed for categorizing agents without resort to such contentious boundaries, rather an agent's category is partially determined by their observed behavior. The determination is partial in the sense that only the probability of category membership can be determined for each agent and usually it is not 0 or 1. In this sense there is only partial determination of class membership, however this is shown to not hinder analysis of behavior of the classes in many dimensions.

Examples of disputed boundaries are not hard to find, determining the poor has probably been the most contentious. The poor are usually identified by specifying some income poverty cut-off and agents with an income level equal to or below that cut-off are in the poor group. There ensued an extensive literature on how the cut-off should be determined with the Sen (1983) versus Townsend (1985) relative versus absolute debate being a feature (Foster, 1998). Townsend advocated a relative measure reflecting a public view of what is socially acceptable (Hill, 2002)[1] usually a proportionate to median income line, for example 50% of median income, Sen advocated an absolute measure reflecting a needs based view, the U.N. $1 and $2 a day measures are an example, Citro and Michael (1995) proposed a combination of the two.

Things are not different when the focus of the analysis is on the middle class. Recently Atkinson and Brandolini (2011) observed that a substantial economics literature has taken a narrow view of the identification of middle class membership in that it has largely been a matter of locating an individual's income in the size distribution of in-

[1]This view is not recent, Adam Smith is often interpreted as having a "relative" sense of necessity when he wrote: "$\cdots$By necessaries I understand, not only the commodities which are indispensably necessary for the support of life, but whatever the custom of the country renders it indecent for creditable people, even the lowest order, to be without" Smith (1976).

comes. For example Easterly (2001) defines the "middle class" as those lying between the 20th and 80th percentile on the consumption distribution. Such a definition fixes the size of the class, it cannot shrink or expand relative to the size of the population so that issues such as the "disappearance or emergence of the middle class" cannot be addressed. Similarly a growing literature on the incomes of the rich class (for example Saez and Veall, 2005) defines the class by some high quantile hence suffering the same problem. Banerjee and Duflo (2008) use an acknowledged *ad hoc* range of values, those households whose daily per capita expenditures valued at purchasing power parity are between \$2 and \$4, and those households between \$6 and \$10 (this can be considered an absolute middle class measure corresponding to an absolute poverty measure). Beach, Chaykovski and Slotsve (1997) consider the proportions of individuals above a scaled up value of median wages using 0.25, 0.5 and 0.75 of the median as frontiers and below a scaled up value of the median wage using 1.25, 1.5 and 1.75 of the median as frontiers (which is akin to a relative middle class measure corresponding to a relative poverty measure).

The most recent disputation as to the value of this type of classification argues that "wellness" in general is a many dimensioned concept so that income of itself is but a vague reflection of societal wellness (Fitoussi, Sen and Stiglitz, 2011). Sen and others (e.g. papers in Grusky and Kanbur, 2006; Kakwani and Silber, 2008; Nussbaum, 2011; Alkire and Foster, 2011) have forcibly argued that limitations to individual's functionings and capabilities should be considered the determining factors in her/his poorness or wellness, again implying that an individual's income will only partially reflect her/his poverty status. With respect to the middle class the analysis has been extended to wealth and property and, largely at the behest of sociologists (see for example Goldthorpe, 2010), it has been extended to occupational status, with control over resources and position in the division of labour being determining features. Indeed this lead Atkinson and Brandolini to argue for the re-integration of different approaches to the concept of middle class. As they observed (p.20) "The entire social stratification has become more complex: ... social class and income distribution largely belong to separate fields of analysis – the former a favorite terrain for sociologists, the latter a topic largely for economists." Unfortunately as more characteristics are added to the list of features that determine class, the boundaries set in any one of them for the determination of class membership inevitably become blurred or at least much more difficult to define., intensifying the arbitrary nature of the process (Alkire and Foster, 2011; Anderson, 2010; Anderson *et al.*, 2006 and 2011). Furthermore many of the determining features of an individuals class, the freedoms they enjoy, the capabilities they possess (as opposed to the extent to which they exercise those capabilities) and the security they experience in their actions are fundamentally unobservable characteristics of an individual agent. However, if these unobservable characteristics do limit or bound observable actions of an individual and if members within each class face similar limits to those characteristics which are different from the limits faced by other classes, it may be possible to discern individual behavior common to a class in the stochastic process that describes their observable actions.

There is an extensive theoretical literature on the size distribution amongst agents defined by a vector $\mathbf{x}$ of outcomes that is the consequence of a stochastic process. Very often the size distribution turns out to be multivariate normal or log normal, Gibrat's law is a classic example of such theorems typically referred to in the statistics literature as Central Limit Theorems (Gibrat, 1930 and 1931; see Sutton, 1997 for a discussion). The power of these laws, like all central limit theorems, is that a (log) normal distribution prevails in the limit almost regardless of the underlying distribution of the stochastic shocks though the mean and variance of the distribution do depend on the parameters governing the process. The choice of normality or log normality under Gibrat's Law is in some sense arbitrary since it depends on the assumed nature of the process of $\mathbf{x}$. If the process is assume to be in levels the resulting distribution of $\mathbf{x}$ will be log normal, if it is assumed to be in terms of $\exp(\mathbf{x})$ the resultant distribution of $\mathbf{x}$ will be normal.

Suppose that the functionings and capabilities set that characterize a particular class (denote it "j") also determine the parameters that govern the stochastic process of an observable vector variable $\mathbf{x}$ for that class. To the extent that the functionings and capabilities of different classes impose different limits on the actions of their members with respect to $\mathbf{x}$, $\mathbf{x}$ at time t will have a particular multivariate distribution $f_j(\mathbf{x})$ that is distinguishable from the corresponding distribution of $f_h(\mathbf{x})$ for class $j \neq h$. Furthermore the distribution of $\mathbf{x}$ in the population will be a mixture of these subclass distributions where the mixing weights are the proportions of society that are members of the respective classes. If these sub distributions and their respective weights can be estimated, much can be said about the behavior and state of wellbeing of classes without resorting to debates about defining boundaries. Indeed it turns out that under certain conditions estimates of the sub distributions yield estimates of the probability that agent i with outcome $\mathbf{x}_i$ is a member of the j'th class $j = 1, \cdots, K$.

In the following, a model that estimates class membership probabilities and simultaneously the influence of possible determinants of class membership is presented. Some possibilities for calculating various descriptive statistics and polarization measures are also suggested. The model is exemplified in section 3 in a study of urban Chinese household incomes drawn from six Provinces over the last decade of the 20th century. We estimated mixture parameters stratified by year along with the effects of individual determinants or factors governing class membership. In addition to these factors the effect of the substantial urbanization that took place over the period together with the impact of the One Child Policy which was introduced in 1978 will be examined. The policy, which was particularly effective in urban China, not only influenced the fertility decisions of households but changed fundamentally the way families were formed in terms of partner choices and investments in children (Anderson and Leo, 2009; 2013). Section 4 concludes.

2 The model.

2.1 Partial definition of group membership.

Assume a finite number K of classes in society whose behaviors are governed by their circumstances to the extent the path of the vector of outcomes $\mathbf{x}$ which describes their behavior follows a distinct process for each of the classes. In this case the joint size distribution of the $\mathbf{x}$ of the j'th group ($j = 1, 2, \cdots, K$), $f_j(\mathbf{x})$ will be distinct from the h'th ($h \neq j$) group distribution $f_h(\mathbf{x})$. With K such groups in a society, the overall distribution $f(\mathbf{x})$ will be a mixture of these distributions (components):

$$f(\mathbf{x}_i) = \sum_{j=1}^{K} w_{ij} f_j(\mathbf{x}_i); \text{ where } \sum_{j=1}^{K} w_{ij} = 1, \ \forall i = 1, \cdots, n. \tag{1}$$

where w_{ij} represents the prior membership probability of agent i to belong to component j.

The propensity of belonging to one or another group is a central issue in the analysis, and to this end the membership probabilities may be modeled as follows. Suppose $\mathbf{z}_i$ is a vector of circumstances or observable characteristics that contribute to determine the membership of agent i with outcome $\mathbf{x}_i$ to class j. Therefore, the effects $\boldsymbol{\beta}_j$ of the observable characteristics $\mathbf{z}$ are related to the probabilities of belonging to a certain component j in the following form:

$$w_{ij} = \text{prob}\left\{\mathbb{I}(C_i = j) = 1|\mathbf{Z}\right\} = g\left(\boldsymbol{\beta}_j, \mathbf{z}_i\right), \quad i = 1 \cdots, n, \text{ and } j = 1, \cdots, K. \tag{2}$$

In this system of equations the w_{ij} sum to one over j and reside in the unit interval so that the link function $g(.)$ will have to satisfy the appropriate constraints much like systems of demand equations which describe expenditure shares. The natural solution would be to use a logistic transform, but a crucial issue is that C_i values are not observed, so that a classical multinomial logistic regression model can not be fitted directly.

The assumptions of our model can be summarized in a system of non-linear equations as follows:

$$\begin{cases} f(\mathbf{x}_i) = \sum_{j=1}^{K} w_{ij} f_j(\mathbf{x}_i) \\ \\ \log\left(\frac{w_{ij}}{w_{i1}}\right) = \mathbf{z}_i \boldsymbol{\beta}_j, \quad j = 2, \ldots, K. \end{cases} \tag{3}$$

where sometimes class-specific parameters can be assumed to be homogeneous, with the exception of a class-specific intercept (Agresti, 2002). An equivalent representation is given by model

$$f(\mathbf{x}_i) = \sum_{j=1}^{K} \frac{e^{\mathbf{z}_i \boldsymbol{\beta}_j}}{\sum_h e^{\mathbf{z}_i \boldsymbol{\beta}_h}} f_j(\mathbf{x}_i),$$

if we let $\boldsymbol{\beta}_1 = \mathbf{0}$ by convention.

Given some assumptions regarding the nature of the $f_j(\mathbf{x}_i)$'s, these components can be specified to belong to some parametric family (joint normality or log normality are popular specifications that can be theoretically rationalized). For example, if the components are assumed to belong to the multi-normal family, the mixture density can be written as:

$$f\left(\mathbf{x}_i;\boldsymbol{\Psi}\right)=\sum_{j=1}^{K} w_{ij} f_j\left(\mathbf{x}_i;\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j\right), \tag{4}$$

where $f_j\left(\mathbf{x}_i;\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j\right)$ denotes the multivariate normal density of the jth component with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$. All the unknown parameters of the mixture model are contained in $\boldsymbol{\Psi}=(\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_K,\boldsymbol{\xi}')'$; in this case $\boldsymbol{\xi}$ consists of the elements of the component means $\boldsymbol{\mu}_1,...,\boldsymbol{\mu}_K$ and the distinct elements of the component-covariance matrices $\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_K$. After estimation of the $\boldsymbol{\beta}$ parameters one may obtain the subject-specific probabilities as

$$w_{ij}=\frac{e^{\mathbf{z}_i\boldsymbol{\beta}_j}}{\sum_h e^{\mathbf{z}_i\boldsymbol{\beta}_h}}.$$

Finite mixture models provide great flexibility in modeling unknown and heterogeneous distributional shapes with modality, skewness and non standard distributional characteristics, whilst retaining some of the analytic advantages of parametric methods. In the past decades, the extent and the potential of the applications of mixture models have widened considerably in many fields[2]. The cost of achieving such flexibility is the increase in the number of parameters to be estimated with the number of components K.

Maximum likelihood (ML) estimation of mixture model parameters is facilitated by use of the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) which has been proposed as an iterative method for solving incomplete data problems. In mixture models, the incompleteness refers to the assignment of each data point to the components of the mixture. Assuming the data $\mathbf{X}=\{\mathbf{x_1},\cdots,\mathbf{x_n}\}$ independently generated from (4), the log-likelihood of the parameters given the data is:

$$\ell\left(\boldsymbol{\Psi}|\mathbf{X}\right)=\sum_{i=1}^{n}\log\left(\sum_{j=1}^{K} f_j\left(\mathbf{x_i}|\mu_j,\Sigma_j\right)\cdot\frac{e^{\mathbf{z}_i\boldsymbol{\beta}_j}}{\sum_h e^{\mathbf{z}_i\boldsymbol{\beta}_h}}\right). \tag{5}$$

The maximum likelihood principle ensures that the best model of the data is the one with parameters that maximize (5). Maximizing (5) is quite difficult numerically because of the sum of terms inside the logarithm. The likelihood function can however be simplified by assuming the existence of additional but missing information. The

[2]See e.g. McLachlan and Peel, 2000, and Melnykov and Maitra, 2010, for detailed and up-to-date reviews into theory and applications of mixture models.

intuition is that if one had access to a latent random variable indicator depending on which data point i was generated by which component j, the "complete data" log-likelihood function would be:

$$\ell_c\left(\boldsymbol{\Psi}|\mathbf{X}, C\right) = \sum_{i=1}^{n}\sum_{j=1}^{K}\mathbb{I}(C_i = j)\log\left(f_j\left(\mathbf{x_i}|\mu_j, \Sigma_j\right)\right) + \sum_{i=1}^{n}\sum_{j=1}^{K}\mathbb{I}(C_i = j)\log(w_{ij}), \quad (6)$$

where $\mathbb{I}(C_i = j)$ represents the indicator variable that takes values 0 or 1 according to whether unit i belongs to component j; and w_{ij} is as defined in (2).

The log-likelihood (6) no longer involves the logarithm of sum but since C is unknown, ℓ_c can not be utilized directly and therefore its expectation is involved in the estimation process. In the EM terminology, $\mathbb{I}(C_i = j)$ is treated as missing information and its posterior expectation estimated in the E-step. Note that *a priori* $\Pr(C_i = j|\mathbf{z}_i) = w_{ij}$. Starting from initial values of the vector of parameters $\boldsymbol{\Psi}^{(0)}$ of $\boldsymbol{\Psi}$ the EM algorithm consists of a sequence of alternate Expectation and Maximization steps until a satisfactory degree of convergence occurs to the ML estimates.[3] The E-step, at the generic iteration $\nu+1$, requires the computation of the conditional expectation of ℓ_c given the data: $\mathbb{E}_{\boldsymbol{\Psi}^{(\nu)}}\left\{\ell_c\left(\boldsymbol{\Psi}|\mathbf{X}, C\right)\right\}$. Since the complete-data log-likelihood ℓ_c is linear in the unobservable data $\mathbb{I}(C_i = j)$, the E-step simply requires the calculation of the subjects' current conditional probabilities of belonging to each component j, τ_{ij}:

$$\mathbb{E}_{\boldsymbol{\Psi}^{(\nu)}}\left\{\mathbb{I}(C_i = j)|\mathbf{X}, \mathbf{Z}_i\right\} = \text{prob}\left\{\mathbb{I}(C_i = j) = 1|\mathbf{X}, \mathbf{Z}_i\right\} \quad (7)$$

that yields the posterior probability that agent i with outcome $\mathbf{x_i}$ belongs to the j-th component of the mixture:

$$\tau_{ij} = \frac{w_{ij}f_j(\mathbf{x_i})}{\sum_{h=1}^{K} w_{ih}f_h(\mathbf{x_i})}, \text{ for } i = 1, \cdots, n \text{ and } j = 1, \cdots, K. \quad (8)$$

For the sample, this is an $n \times K$ matrix with its rows adding up to unity. In the M-step, we estimate the parameters of each distribution. In estimating the parameters for component j, the indicator of belonging to component j is replaced by its

[3] It is well known that the likelihood function of normal mixtures is unbounded and the global maximizer does not exist (McLachlan and Peel, 2000). Therefore, the maximum likelihood estimator of $\boldsymbol{\Psi}$ should be the root of the likelihood equation corresponding to the largest of the local maxima located. The solution usually adopted is to apply a range of starting solutions for the iterations.

conditional expectation, equal to the subject's conditional probability of belonging to the component. The latent parameters $\boldsymbol{\beta}$ are updated through a classical Newton-Raphson iteration for multinomial logistic models.[4] A final issue concerns estimation of standard errors of the parameters involved. The EM algorithm does not yield these as a by-product. We proceed here using Oakes' identity (Oakes, 1999), which requires additional computational effort. According to Oakes' identity, minus the expected information matrix is given by the second derivative of the conditional expected value of the complete data log-likelihood given the observed data, plus the first derivative of the score, for the same expected log-likelihood, with respect to the current value of the parameters. This component is directly obtained at the M step of the EM algorithm, while the second can also be obtained in closed form similarly.[5] The standard errors are then estimated as the square root of the diagonal of the expected information matrix.

Estimation of the posterior probabilities τ_{ij} provides K group membership indices for each agent in the population. How well determined are the respective groups? Effectively it is possible for the group distributions to overlap, that is for agent i with outcome $\mathbf{x_i}$ to potentially be a member of more than one group. To the extent that these distributions do not overlap (perfect segmentation in the terminology of Yitzaki, 1994) knowing an individual's vector of outcomes will completely determine an agent's group and all of the agents in a group. To the extent that they do overlap an agent's vector of outcomes will only partially define her group member ship in the sense that the probability of her being in a particular group is all that can be obtained. Trends in the extent of overlap between class distributions reflect the extent to which groups are polarizing or converging (Anderson, Linton and Wang, 2012). Unfortunately when there is little overlap, as will be found to be the case here for some groups, the overlap measure is unreliable so resort must be made to the trapezoidal measure (which is asymptotically normal) given by $1/2\left(f(\mathbf{x}_{m,j}) + f(\mathbf{x}_{m,h})\right)||\mathbf{x}_{m,j} - \mathbf{x}_{m,h}||$ where $\mathbf{x}_{m,j}$ is the modal vector of the j-th group and $f(\mathbf{x}_{m,j})$ is the corresponding value of the pdf at the modal point (Anderson, 2010; Anderson, Linton and Leo, 2012).

What of other characteristics of agents in a particular group? The probability

[4] It can be noted that when $\mathbf{z}_i$ is a column-vector of ones, that is, we do not include covariates but only an intercept, $w_{ij} = w_j$, $\forall i$ and a closed form solution exists for class weights w_j, the marginal probabilities.

[5] Computational details are reported in Appendix A.2.

measures can serve as selectors which permit the calculation of a whole range of class characteristics. Suppose agent i with $\mathbf{x}_i$ reports the status of a characteristic z (suppose for example it is an education, health or family background index) as z_i, then a whole range of indices for the status of the j'th class with respect to z can be calculated. For example means, variances and Foster-Greer-Thorbecke (FGT, 1984) M-generalized measure of poverty with respect to a cutoff z^* would respectively be:

$$\overline{z_j} = \frac{\sum_{i=1}^{n} \tau_{ij}\, z_i}{\sum_{i=1}^{n} \tau_{ij}}$$

$$\mathrm{V}(z_j) = \frac{\sum_{i=1}^{n} \tau_{ij} \left(z_i - \overline{z_j}\right)^2}{\sum_{i=1}^{n} \tau_{ij}}$$

$$\mathrm{FGT_M}(z_j) = \frac{\sum_{i=1}^{n} \tau_{ij} \mathbb{I}\left(z^* - z_i\right) \left(\frac{z^* - z_i}{z^*}\right)^{\mathrm{M}-1}}{\sum_{i=1}^{n} \tau_{ij}};$$

where

$$\mathbb{I}\left(z^* - z_i\right) = 1 \;\text{ if }\; z^* - z_i > 0 \;\text{ else } = 0.$$

Naturally these statistics provide instruments for making interclass comparisons so that various between class distance and dominance statistics may be computed addressing such questions as how much better off are the middle class than the poor in the dimension of z, or how polarized are particular classes. M'th order dominance comparisons between group j and group h can be made by considering the incomplete subgroup moments:

$$\mathrm{F}_j(z^*) - \mathrm{F}_h(z^*) = \sum_{i=1}^{n} \left[\left(\frac{\tau_{ij}}{\sum_{i=1}^{n} \tau_{ij}} - \frac{\tau_{ih}}{\sum_{i=1}^{n} \tau_{ih}} \right) \left(z^* - z_i\right)^{\mathrm{M}-1} \mathbb{I}\left(z_i \leq z^*\right) \right].$$

3 Evolution of the income classes in urban China.

3.1 Data issues.

During the last part of the last century, urban Chinese households experienced profound changes (e.g. Tao Yang, 1999; Wu and Perloff, 2005). The One Child Policy intervention introduced in the late 1970's changed fundamentally the nature of the family in many respects in subsequent years. The Economic Reforms, also instigated in the late 1970's, appeared to promote unprecedented growth in urban incomes (as reported in Ravaillon and Chen, 2007), the average annual growth rate of urban household incomes per person over the period 1981-2002 was over 6%). In addition there was a massive migration to the cities (in 1981 less than 20% of the Chinese population was urbanized, by 2002 almost 40% of a growing population was urbanized). 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 grouped and evolved into classes.

Micro-data on ten cross-sectional annual surveys of urban households from three coastal and three interior provinces in China from 1992 to 2001, coming from Urban Household Surveys, are used to study the evolution of household income classes and their determining characteristics over the period.

China is one of the few countries in which rural and urban household surveys were separately implemented. The Urban Household Survey (UHS), promoted by the Chinese National Bureau of Statistics (NBS), is a national survey that collects individual and households data using a questionnaire and sampling frame designed to investigate the phenomena of urban unemployment and poverty in China[6].

The six selected provinces are characterized by different levels of per capita GDP: Sichuan with per capita GDP in the year 2010 equal to 3,104 US $ (source: China's statistics yearbook), Shaanxi (per capita GDP equal to 3,996 US$), Hubei (per capita GDP equal 4,079 US$), Jilin (per capita GDP equal to 4,614 US$), Shandong (per capita GDP equal to 6,078 US$), Guandong (per capita GDP equal to 6,440 US$). In each year the average sample in these provinces is around 4000 households resident in 13 cities. The main content of the UHS, besides demographic characteristics, includes the basic conditions, such as living expenditures for consumption, purchase of major commodities, durable consumer goods owned at the end of the year, housing conditions and cash income and expenditures.

Analysis of class membership is based on the household disposable income from all sources, that is the total of the personal income of all the members of the family. The analysis is carried out on household income adjusted for different household sizes using

[6] As described in Fang *et al.* (1998), surveys of urban households started in 1956, were suspended from 1966 to 1979, and resumed in 1980. In 1984, the Urban Social and Economic Survey Organization was set up. The corresponding survey teams for urban surveys were established in 30 provinces. The number of urban households surveyed increased from 8,715 in 1981 to around 33,000 in 1987 and has remained about the same until the 2000's. On the quality of household income surveys in China see Bramall (2001).

the square root scale, a scale which divides household income by the square root of household size. Household incomes are reported in 1994 prices using the corresponding national deflator. Since regional price differences are expected to be wide in China (Brandt and Holz, 2006), comparison across provinces can not be implemented without employing spatial price indices (SPI) for adjusting spatial price differentials. We used Gong and Meng (2008) SPI as regional deflators. They derived SPI for different provinces for urban China during the period 1986-2001 using the Engel's curve approach (Hamilton, 2001), which overcome some problems suffered by the most commonly used basket cost method.[7]

3.2 Model Estimation.

3.2.1 Parameters of mixture distributions and polarization cohesion.

Model (3) was estimated using size adjusted household incomes as the observable outcome assuming Gaussian component densities.[8] Given our data structure, estimates were carried out pooling observations from all years but allowing component parameters of the mixture to vary over time.[9] The procedure is robust enough to prevent the emergence of bizarre components, like very flat components with large dispersion and very small marginal probabilities or components with similar means but different shapes due to their disparate variances. However, it is also flexible enough to allow for temporal dynamics since it is possible to estimate mixture parameters $\theta_{jt} = (\mu_{jt}, \sigma_{jt})$ stratified by year.

Given sample size of between 4 and 5 thousand observations per year, the number of components has been assessed by using the Bayesian's Information Criteria

[7]It should be emphasized that the authors find, as expected, that high income provinces such as Guandong are ranked as high price provinces, while low income provinces as Shaanxi have low prices. They also find similar SPI variation over time across different methods, especially between Engel's curve approach and the basket cost method employed by Brandt and Holz (2006).

[8]The assumption of normality may be too restrictive, since in principle any functional form can be taken into account. The choice of normality stems from a threefold motivation. Firstly, mixture of normal distributions form a much more general class. In fact, any absolutely continuous distribution can be approximated by a finite mixture of normals with arbitrary precision (Marron and Wand, 1992). Secondly, a mixture model of normals seems to capture better than other functional forms the idea of a polarized economy where relatively homogeneous groups of households are clustered around their expected incomes. Thirdly, assumption of normality, following Gibrat's law, results from additive shocks to the expected income of each strata.

[9]All the computing was carried out in R (2012). To avoid numerical underflow, we used a log-scale summation device (Farcomeni, 2012). Functions are available on request.

Table 1: *The choice of the number of components according to the Bayesian Information Criterion (BIC).*

k	log-lik	# parameters	BIC
2	-350313	73	701402
3	-340529	126	682398
4	-339571	179	681046
5	-339292	232	681052

(BIC), which is known to work well with large sample size. BIC adds a term in the log likelihood so to penalizes the extra complexity of the model and in this context is particularly helpful in finding a parsimonious parametrization. Although regularity conditions do not hold for mixture models, Keribin (2000) showed that BIC is consistent for choosing the number of components in a mixture. According to BIC, a four-component mixture seems to be the 'best' parsimonious model (see Table 1). Adding a fifth component to the mixture does not improve the fit of the model enough to warrant the additional complexity. Our choice of four-component mixture is corroborated by the fact that each component is characterized by distinct means, relatively modest dispersion and not negligible size and adding components can play a role in improving the fit of the whole distribution but may be unacceptable in terms of economic interpretability (Longford, 2008). As a matter of fact, the four-component solution facilitates the economic determination of each component. They can be interpreted as "poor", "lower-middle", "upper-middle" and "rich" income groups.

Table 2 provides a summary of the model fits stratified by years: the estimated mean (μ_{jt}) and standard deviation (σ_{jt}) of each normal component along with its corresponding mixing proportion ($\overline{w}_{jt}$). Figures 1 visually compare the fitted four component mixtures for some years of the analysis with the corresponding estimated kernel density.[10]

Table 2 reveals real annual income growth rates of 3.53% for the poor, 5.21% and 6.13% for the lower and upper middle class respectively, and 7.25% for the rich over the

[10]For the purpose of comparison, the variance of each component population was inflated by a factor of $1 + h^2/\sigma_i^2$ to match that of the kernel density, where the kernel is Gaussian and h its estimated bandwidth.

Table 2: *Estimated parameters of the mixture components for each year.*

	Poor			Lower-middle			Upper-middle			Rich		
Year	μ	σ	$\overline{w}$	μ	σ	$\overline{w}$	μ	σ	$\overline{w}$	μ	σ	$\overline{w}$
1992	852	220	10.3%	1345	289	35.1%	1971	405	41.0%	3214	568	13.6%
1993	856	228	10.7%	1441	326	34.0%	2302	513	41.0%	3796	766	14.3%
1994	877	248	11.0%	1525	371	33.0%	2478	589	41.5%	4106	838	14.5%
1995	955	240	11.4%	1622	363	31.9%	2533	581	41.7%	4019	799	15.0%
1996	1080	283	11.8%	1784	434	32.1%	2575	658	41.3%	4172	899	14.8%
1997	1038	283	11.6%	1741	407	32.9%	2649	625	41.5%	4207	858	14.0%
1998	1058	286	10.9%	1791	455	32.1%	2804	699	41.7%	4630	1043	15.2%
1999	1154	375	10.0%	1941	478	31.6%	3078	755	42.0%	5152	1094	16.3%
2000	1128	380	10.4%	1994	531	31.8%	3197	843	40.4%	5520	1349	17.5%
2001	1164	374	9.8%	2124	588	32.6%	3368	942	39.3%	5924	1659	18.3%

Note: μ and σ are in 1994 constant yuan, $\overline{w}$ the mean of individuals prior membership probabilities.

period. The marginal probability of being in the poor group slightly increases at the beginning of the period (moving from 10.3% in 1992 to 11.8% in 1996) and decreases since 1996 reaching a level of 9.8% in 2001. The lower-middle income group becomes stable stable in size after the first two years and similarly the upper-middle group. The rich group mixing proportions tend to increase especially in the last three years reaching the level of 18.3% in 2001. The within group inequalities grow for all groups (which is consistent with Gibrat's law).

The size of poor group determined by the mixture model can be compared with the traditional (absolute and relative) identification of the poor by employing a poverty cut-off. Following Ravaillon and Chen (2007) we used an absolute poverty line of 1200 Yuan for urban areas at 2002 prices, that we converted at 1994 prices for congruity with our income data. We also fixed a relative poverty line at 60% of the median income for each year. The corresponding head count ratios are reported in Table 3.

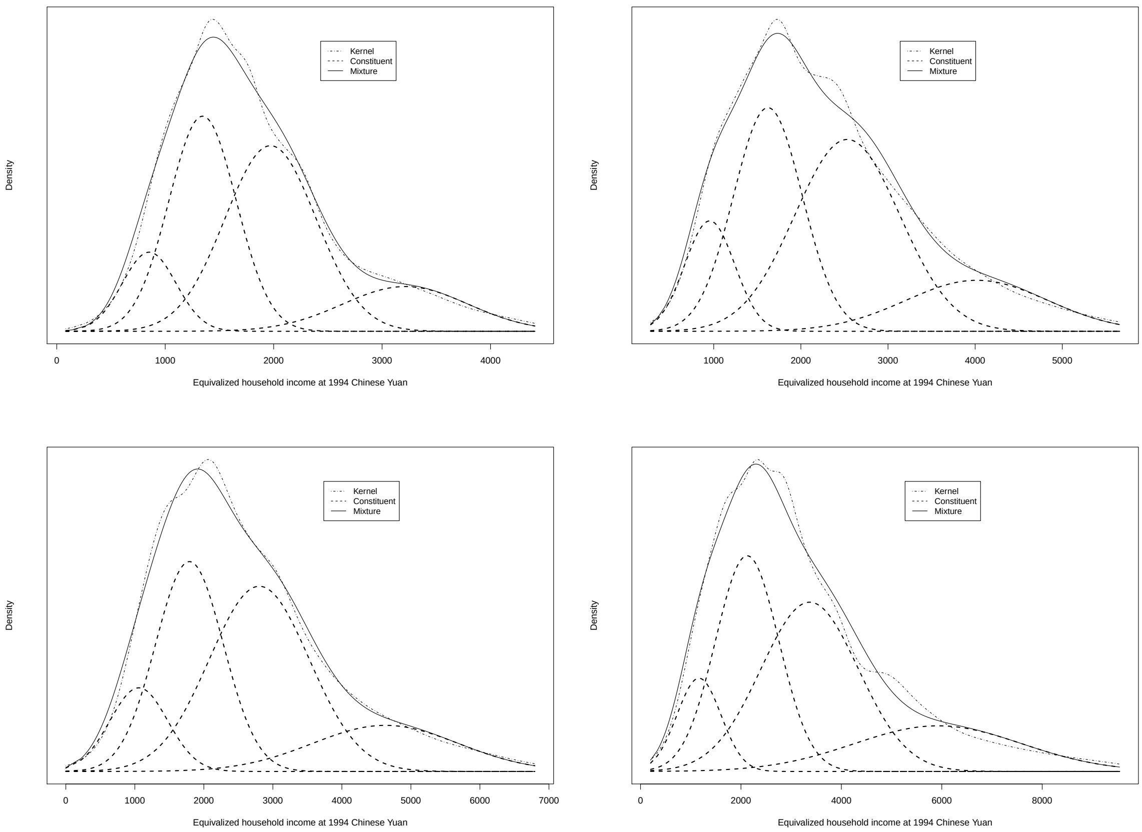
The size of the poor group estimated by the mixture model lies between the absolute poverty measure and the relative one. While the absolute poverty drastically decreases over time, the percentage of households below the relative poverty line exhibits a fluctuating pattern until 1997 and it increases afterwards, in contrast with the estimated size of the poor group according to the model.

As previously described, in the mixture approach the membership of each class is not determined with certainty, but each household has attached an estimated proba-

Table 3: *Absolute and relative measure of poverty (head count index in percentage) compared with the size of the poor component in the mixture model*

Year	Absolute Poverty	Relative Poverty	Size Poor Group
1992	9.3	12.2	10.3
1993	9.2	16.1	10.7
1994	8.5	17.4	11.0
1995	6.2	16.0	11.4
1996	4.6	14.6	11.8
1997	5.0	15.5	11.6
1998	5.0	16.4	10.9
1999	3.4	16.6	10.0
2000	4.2	17.8	10.4
2001	3.5	18.3	9.8

Figure 1: *Kernel density estimation and the (inflated) four-components mixture model fit, 1992, 1995, 1998 and 2001.*



bility of belonging to each component of the mixture (see eq. 8). As the components overlap, there is considerable uncertainty about the household's allocation, while if the components are well separated, the conditional probabilities τ_{ij} tend to define a partition/segmentation of the population.

Perusal of figure 1 will indicate very little overlap between rich and poor and rich and lower middle classes so recourse is made to the trapezoidal measures (Anderson, Linton and Leo, 2012) reported in Table 4. As can be seen, given the asymptotic normality and standard deviations of the estimates and by taking differences in the trapezoidal measures, there is very strong evidence of polarization between all of the classes over the period with some minor retrenchments over the 1995, 1996 years.

Table 4: *Mixture trapezoid measures. Standard errors in parenthesis.*

year	R-UM	R-LM	R-P	UM-LM	UM-P	LM-P
1992	22.724	37.573	51.534	13.550	11.517	12.415
	(0.048)	(0.054)	(0.077)	(0.035)	(0.058)	(0.053)
1993	23.925	42.990	60.027	17.095	12.880	14.191
	(0.048)	(0.057)	(0.083)	(0.037)	(0.061)	(0.054)
1994	24.599	44.514	63.150	17.702	13.534	14.919
	(0.047)	(0.056)	(0.082)	(0.037)	(0.061)	(0.054)
1995	22.784	42.011	61.073	17.077	14.108	15.571
	(0.045)	(0.055)	(0.080)	(0.037)	(0.061)	(0.054)
1996	23.043	38.752	57.233	13.725	13.822	15.088
	(0.046)	(0.054)	(0.078)	(0.037)	(0.059)	(0.053)
1997	23.041	41.175	59.156	16.223	13.945	15.287
	(0.047)	(0.055)	(0.078)	(0.036)	(0.058)	(0.052)
1998	25.055	44.083	64.194	17.116	14.176	15.500
	(0.046)	(0.055)	(0.084)	(0.037)	(0.062)	(0.055)
1999	27.564	48.661	65.293	18.628	13.820	15.287
	(0.045)	(0.054)	(0.078)	(0.037)	(0.060)	(0.054)
2000	28.575	49.672	68.794	18.678	14.811	16.358
	(0.045)	(0.055)	(0.082)	(0.037)	(0.061)	(0.054)
2001	29.129	49.869	72.408	18.318	16.141	17.799
	(0.046)	(0.056)	(0.090)	(0.037)	(0.066)	(0.057)

3.2.2 Determinant effects

With poor as the baseline category, Table 5 contains EM estimates from model (3) of effect parameters $\boldsymbol{\beta}_j$, along with their standard errors and relative p-values. The sample refers to period 1992–2001 and explanatory variables include: demographics (age and age squared, household size), employment status of household head (recoded as: State-owned enterprizes (SOE) employee (reference group), employee of collective enterprizes, other employee and self-employed, employed after retirement, retired, others not working), education of the household head (coded 1-7, from 1 (no schooling) to 7 (university graduates)[11]), family status (recoded as single parent or not). We also added province of residence (Shaanxi as reference), a time trend, and the urbanization index of the province of each year.[12]

In order to investigate possible different effects due to the one-child policy (OCP), we separate out the sub-population that most likely was not involved in this limitation (i.e. people who were at least 39 year old in 1978 when the policy was introduced). The pre one-child policy sub-population represents around 17% of our sample.

Consider a baseline Chinese household (a medium size with household head being a single parent, born before the One Child Policy, residing in a medium-sized city in Shaanxi province, of average age, average education and state employees), the prediction equation for the log odds of this household of being in the middle income group instead of poor group is:

$$\log(\frac{\hat{\pi}_{lower-middle}}{\hat{\pi}_{poor}}) = 6.310 - 0.006 \times t + 0.069 \times \overline{age} + 0.480 \times \overline{\mathrm{E}ducation}$$
$$-4.84 \times \overline{\mathrm{U}rbanization\ Index} - 1.270 \times \overline{\mathrm{H}ousehold\ size}$$

There are strong provincial factors both in levels and trends. Holding other effects constant, the odds for households who live in Jilin of being in the rich group rather than in the poor are 0.60 times (40% lower than) the odds for a family who lives in Shaanxi. The odds for families living in the other provinces of belonging to the rich group are higher than the odds for residents in Shaanxi. The provincial trends are for increasing odds for families being in the rich group (*versus* poor group) except for Shandong and particularly evident for Guangdong (see also Figure 2).

With respect to provincial urbanization rates, holding other effects constant, the odds of Chinese families of belonging to the other groups rather than to the poor are statistically significant and important in size and even more so when compounded for pre-OCP families. The strong negative urbanization effect for lower middle, upper middle and rich classes reflects the fact that a concomitant of a families transit from rural to urban China was entry into the poor class initially swelling its numbers relative to the other classes.

[11] We decided to treat the categorical variable of education as a continuous variable in our final model since the estimated effects of education are almost perfectly linear.

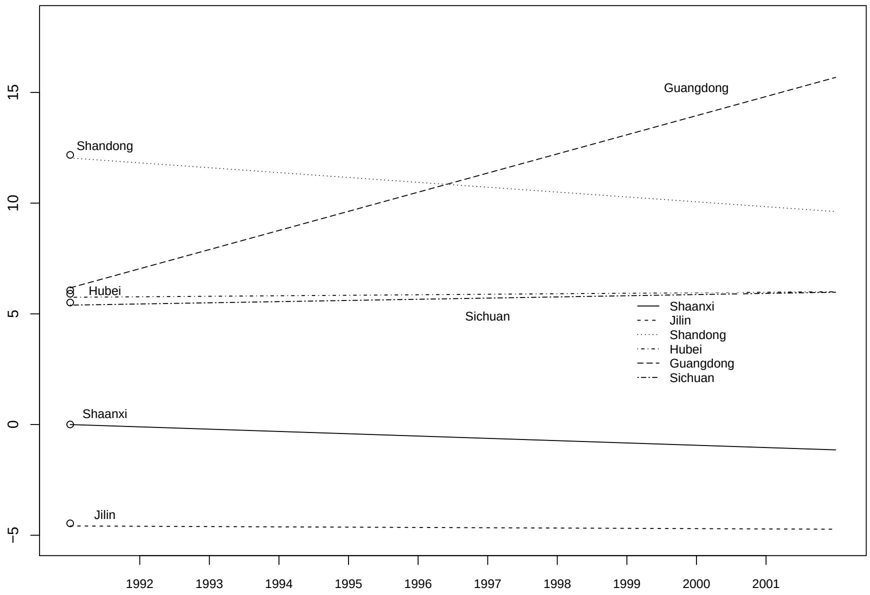
[12] See Appendix A.1 for details on its construction.

Table 5: *Estimated parameters of effects on the odds that chinese households belong to the lower middle, upper midlle and rich group. Poor group is set as reference category.*

	LOWER-MIDDLE					UPPER-MIDDLE					RICH				
	Estimates	s.e.	t val.	$P(>\|t\|)$		Estimates	s.e.	t val.	$P(>\|t\|)$		Estimates	s.e.	t val.	$P(>\|t\|)$	
Shaanxi	-														
Jilin	-0.506	0.070	-7.183	0.000	***	-3.074	0.076	-40.577	0.000	***	-4.580	0.136	-33.716	0.000	***
Shandong	10.657	0.167	63.801	0.000	***	11.571	0.145	79.726	0.000	***	12.038	0.202	59.558	0.000	***
Hubei	5.099	0.102	50.127	0.000	***	5.730	0.089	64.110	0.000	***	5.748	0.125	45.800	0.000	***
Guangdong	4.999	0.133	37.540	0.000	***	5.574	0.119	46.790	0.000	***	6.174	0.169	36.584	0.000	***
Sichuan	5.574	0.111	49.996	0.000	***	5.510	0.098	56.133	0.000	***	5.393	0.139	38.925	0.000	***
time-trend	-0.006	0.009	-0.654	0.513		-0.063	0.008	-7.944	0.000	***	-0.104	0.011	-9.339	0.000	***
Shaanxi:trend	-									***					
Jilin:trend	0.000	0.011	-0.009	0.993		0.098	0.011	8.546	0.000	***	0.091	0.020	4.493	0.000	***
Shandong:trend	-0.159	0.012	-13.331	0.000	***	-0.161	0.010	-15.463	0.000	***	-0.116	0.015	-7.903	0.000	***
Hubei:trend	0.141	0.012	11.939	0.000	***	0.058	0.011	5.457	0.000	***	0.127	0.015	8.490	0.000	***
Guangdong:trend	0.799	0.013	60.056	0.000	***	0.851	0.011	74.992	0.000	***	0.968	0.015	63.291	0.000	***
Sichuan:trend	0.091	0.011	8.075	0.000	***	0.091	0.010	8.865	0.000	***	0.157	0.014	10.848	0.000	***
age	0.069	0.007	9.988	0.000	***	0.209	0.006	32.411	0.000	***	0.351	0.009	37.991	0.000	***
age2	0.000	0.000	-3.537	0.000	***	-0.001	0.000	-18.217	0.000	***	-0.002	0.000	-24.225	0.000	***
Education	0.549	0.007	74.771	0.000	***	1.028	0.007	157.074	0.000	***	1.338	0.010	136.452	0.000	***
Household size	-1.567	0.015	-103.897	0.000	***	-3.218	0.017	-189.133	0.000	***	-5.226	0.034	-151.717	0.000	***
Urbanization index	-4.273	0.087	-49.271	0.000	***	-4.370	0.075	-58.017	0.000	***	-4.764	0.104	-45.853	0.000	***
Single parent dummy	-0.842	0.054	-15.651	0.000	***	-1.845	0.051	-36.206	0.000	***	-2.783	0.072	-38.738	0.000	***
SOE employee	-									***					
Employee of collective enterp	-0.839	0.027	-31.223	0.000	***	-1.522	0.029	-52.631	0.000	***	-2.399	0.060	-39.935	0.000	***
Employee other types, Self Employed, etc	-0.565	0.040	-14.095	0.000	***	-0.801	0.042	-19.197	0.000	***	-0.759	0.064	-11.824	0.000	***
Retired	0.052	0.044	1.172	0.241		-0.006	0.042	-0.152	0.879		-0.152	0.060	-2.537	0.011	**
House workers, unemp, students	-2.169	0.096	-22.600	0.000	***	-3.797	0.135	-28.183	0.000	***	-4.767	0.247	-19.270	0.000	***
Employed after retirement	2.888	0.135	21.408	0.000	***	4.422	0.094	47.131	0.000	***	4.956	0.116	42.610	0.000	***
Pre OCP dummy variable (dvar)	-0.888	0.141	-6.289	0.000	***	-0.188	0.131	-1.436	0.151	**	-0.306	0.196	-1.562	0.118	
Education:dvar	-0.069	0.017	-3.974	0.000	***	-0.193	0.015	-13.275	0.000	***	-0.091	0.019	-4.691	0.000	***
Household size:dvar	0.294	0.027	10.846	0.000	***	0.369	0.030	12.460	0.000	***	0.416	0.058	7.136	0.000	***
Urbanization index:dvar	-0.569	0.040	-14.132	0.000	***	-0.824	0.035	-23.672	0.000	***	-0.892	0.046	-19.248	0.000	***
Single parent dummy:dvar	1.887	0.097	19.492	0.000	***	2.348	0.098	24.065	0.000	***	3.075	0.143	21.478	0.000	***
SOE employee:dvar	-														
Employee of collective enterp.:dvar	1.719	0.182	9.440	0.000	***	1.220	0.213	5.722	0.000	***	1.803	0.328	5.501	0.000	***
Employee other types, Self Employed:dvar	0.566	0.268	2.110	0.035	*	-0.509	0.288	-1.767	0.077	**	-3.141	0.569	-5.524	0.000	***
Retired:dvar	0.203	0.078	2.591	0.010	**	-0.191	0.070	-2.729	0.006	***	-0.550	0.095	-5.794	0.000	***
House workers, unemp, students:dvar	0.014	0.192	0.072	0.943		-1.349	0.245	-5.502	0.000	***	-2.996	0.420	-7.131	0.000	***
Employed after retirement:dvar	-1.121	0.173	-6.474	0.000	***	-1.836	0.123	-14.866	0.000	***	-2.257	0.152	-14.811	0.000	***
Intercept	5.27	0.173	30.500	0.000	***	5.43	0.164	33.200	0.000	***	5.38	0.243	22.100	0.000	***

*** significant at 1%; **significant at 5%; *significant at 10%.

Figure 2: *Estimated provincial trends of odds of belonging to "rich" component vs. poor component.*



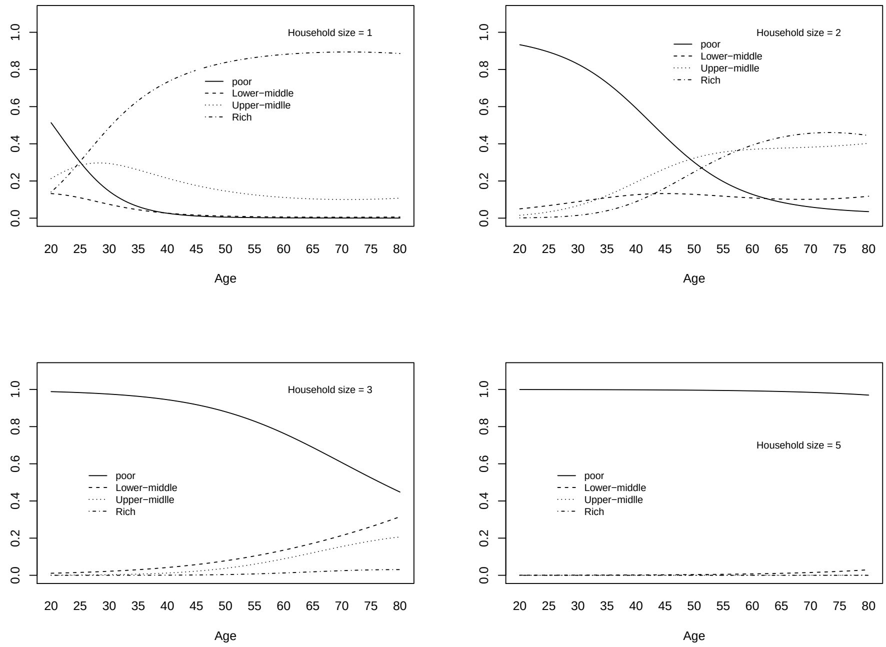
From Table 5 we can estimate the probability of an agent with profile $\mathbf{z}$ to belong to group j as

$$\pi_j = \frac{\exp\left(\boldsymbol{\beta}_j'\mathbf{z}\right)}{1 + \sum_{h=1}^{4} \exp\left(\boldsymbol{\beta}_j\mathbf{z}\right)}$$

with $j = 1, 2, 3, 4$ corresponding to poor, lower middle, upper middle, rich, respectively.

Figure 3 plots the estimated probabilities for a baseline Chinese household of belonging to the poor, the lower middle, the upper middle and the rich group as a function of household size and age of household head. In single-component families the probability of being poor is around 50% for young people but decreases as age increases, and the probability of belonging to the rich group is an increasing function of maturity and reaches the level of around 90% when the household head is around 55 years old. In two-component families the probability of being poor is quite high for young people and starts to decrease only after the age of approximately 35 years. A different pattern is for three-component households: the probability of being in the rich group is essentially zero regardless the age of the household head while the probability of being

Figure 3: *Age profiles: estimated probabilities of belonging to the four groups.*



poor is quite high for almost the entire life. The chance to escape from poor state is given by the possibility to move to the lower and/or upper middle group but this is possible only after the age of 60. As evident form the right bottom panel of Figure 3, when the size of the family is five[13] the age profiles are much flatter indicating that the probability of being poor or rich is not correlated with household vintage. These results confirm that the odds of a family being rich are always lower than the odds of being poor as household size increases. For example, each additional member in the household reduces the odds of being in the lower middle group by 0.209, i.e. decreases them by around 80%.

Holding constant the other predictors, the odds of being in the rich group rather than in the poor for a reference family whose head was born in the pre-OCP period are approximately 26% lower than a similar family whose head was born in post-OCP period. Single parent family status results in a significantly increased risk of poverty group membership for pre-OCP status families but the risk is lower for post-OCP families.

[13]Results are almost identical for families with size equal or greater than four.

Holding other effects constant, a one-level increase in education multiplies the odds of being in the lower-middle group rather than in the poor by 1.73, i.e. increases them by 73%. A one-level increase in education instead multiplies the odds of being wealthy (upper-middle group) by approximately 180%. Group membership probabilities as a function of education are plotted in Figure 5. Poor group membership probabilities strongly diminish with high levels of education with the effects diluted for post as opposed to pre OCP families in the poor group. On the other hand, high levels of education increase the probability to belong to the middle or rich group but do not affect significantly the probability of being in the lower middle group. These effects in the upper-middle group are more pronounced for the post OCP families, while effects are more evident in the rich group for the pre OCP families.

For post-OCP families, if the household head works in collective enterprises the odds of belonging to the lower middle group rather then to the poor are lower than (of approximately 56%) the odds for households whose head is employed in state owned enterprises. The odds of belonging to the rich group rather than to the poor group are 90% lower than the odds of SOE employees. Similarly, the odds for households whose head is self-employed or employed in other type of enterprises of belonging to the lower middle group rather than to the poor group are about 40% lower than the odds of households whose head is employed in a state enterprise. The odds of belonging to the rich group rather than to the poor group are 53% lower than the odds of SOE employees. Being unemployed, houseworker or student largely reduces the odds of being in higher level income groups than in the poor group.

These effects present important differences for pre-OCP families, at least in some cases. Largely speaking opposite effects are observed in the lower-middle group for household head employed in collective enterprises. For pre-OCP families the odds of belonging to the lower-middle income group rather than to the poor group are 140% higher than the odds for families whose head is SOE employed. If we look at pre-OCP families whose head is retired, the odds of being rich rather than poor are 50% lower than someone who is SOE employed, but the odds extremely increase when household heads are employed after retirement. The estimated odds (relative to SOE employees) are shown in Figure 4. Poor group is the reference category.

Figure 4: *Estimated log odds ratios with relative ±2 standard erros for occupation categories: post-OCP families (full dots) and pre-OCP families (stars). Poor group is the reference category.*

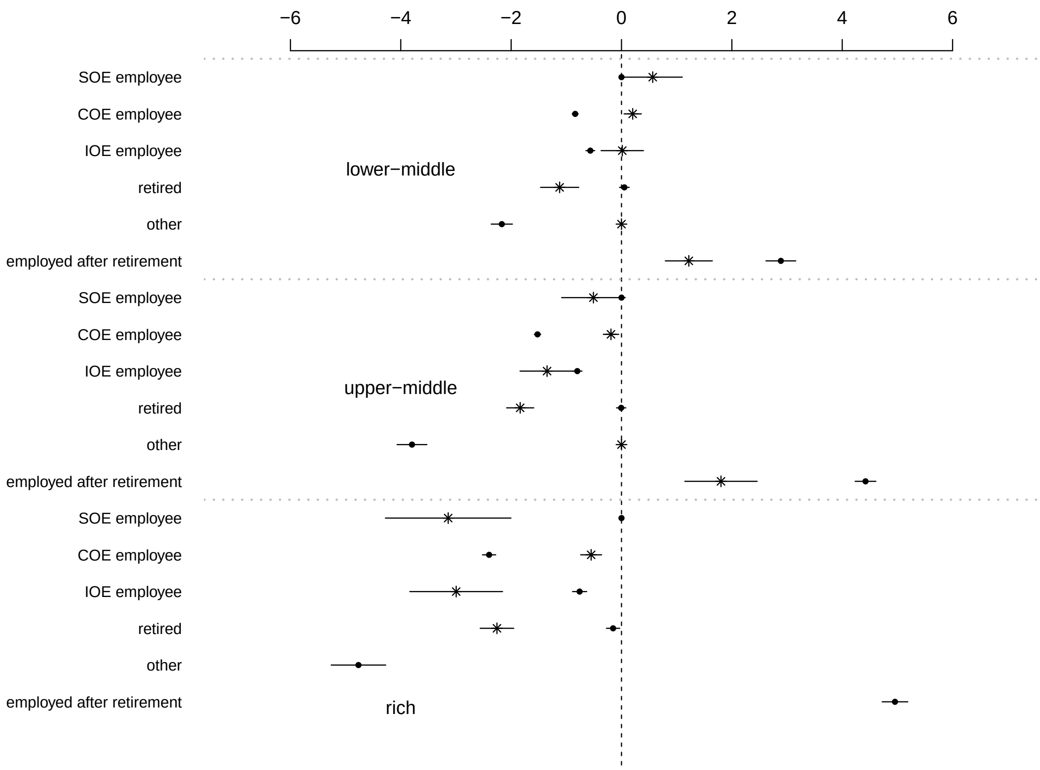
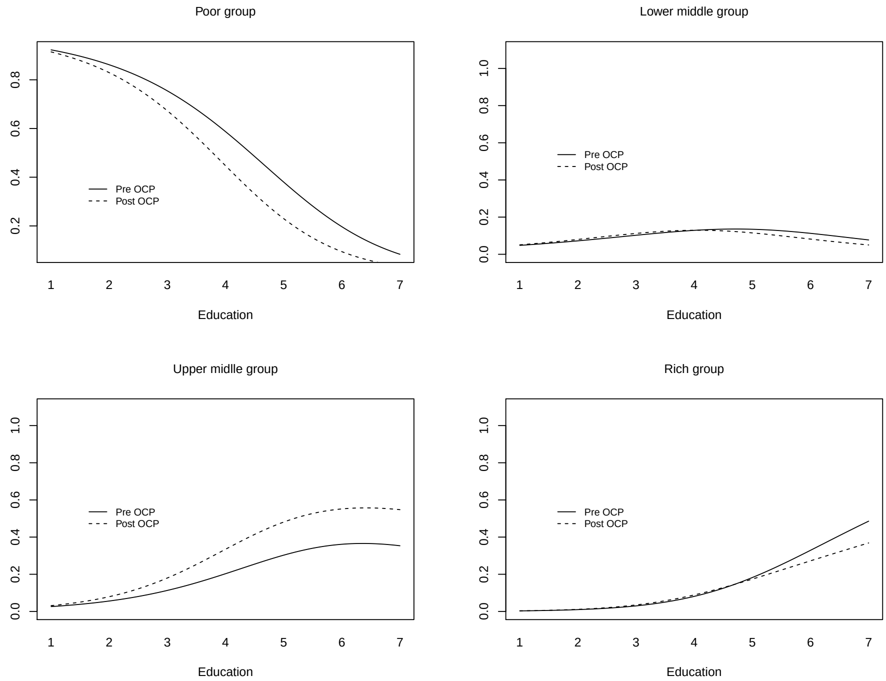


Figure 5: *Education profiles - estimated probabilities of belonging to the four groups for Pre and Post OCP families.*



4 Concluding remarks.

There has been a long standing tradition in economics of classifying agents into groups in order to study their collective wellbeing, the poor, middle and rich classes being cases in point. Usually this involves the defining an artificial boundary or frontier to establish class membership. Here a new technique have been proposed for partially determining class status without resort to artificial boundary assumptions. In essence the classes are determined by similarities in the behavior of agents (households) with respect to economic and social variables. In the present context classes are defined by commonality in size distribution of household vector of outcomes which is modeled as a finite mixture of sub-distributions where each sub-distribution corresponds to the size distribution for a particular class. Class membership determination is partial in the sense that only the probability of the class status of a particular household can be determined. However this facilitates study of trends in the size and summary statistics of the respective classes together with the factors that influence the probability of class status and hence class membership of individual households. As a substantive illustration, restricted to a univariate outcome, this technique has been applied in a study on disposable size-adjusted income of urban households in six Chinese provinces (three coastal and three interior) over the last decade of the 20th Century from 1992 to 2001. This was a period during which urban China was experiencing rapid growth, both economically and in terms of a population flight from the land. Over the sample period four classes were determined which, for want of better terminology, were named Poor, Lower Middle, Upper Middle and Rich classes.

All classes enjoyed income growth throughout the period with lower classes systematically growing more slowly than upper classes. As a consequence the respective classes were all clearly polarizing over the period with respect to each other. Class membership was significantly influenced by all the factors considered in a predictable fashion.In particular the rapid rate of urbanization, the one child policy and economic growth enjoyed by the family had profound effects upon class membership probabilities.

References

Agresti, A. (2002) *Categorical Data Analysis*, Wiley, New York.

Alkire, S. and J.E. Foster (2011). “Understandings and misunderstandings of multidimensional poverty measurement”, *Journal of Economic Inequality*, 9, 289-314.

Anderson G.J. (2010). “Polarization of the Poor: Multivariate Relative Poverty Measurement Sans Frontiers”, *Review of Income and Wealth*, 56(1), 84–101.

Anderson G.J., I. Crawford and A. Leicester (2006). “Statistical Tests for Multidimensional Poverty Analysis”, in N. Kakwani and J. Silber (eds.), *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave Macmillan.

Anderson G.J. and Leo T.W. (2009). “Child Poverty, Investment in Children and Generational Mobility: The Short and Long Term Wellbeing of Children in Urban China After The One Child Policy”, *Review of Income and Wealth*, 55, 607–629.

Anderson G.J., Linton O. and Leo T.W. (2012). “A Polarization-Cohesion Perspective on Cross Country Convergence”, *Journal of Economic Growth* 17, 49–69.

Anderson, G., Linton, O., and Whang, Y.J. (2012). ”Nonparametric Estimation and Inference about the Overlap of Two Distributions”, *Journal of Econometrics*, 171(1), 1–23.

Anderson G.J., and Leo T.W. (2013). "An Empirical Examination of Matching Theories: The One Child Policy, Partner Choice and Matching Intensity in Urban China", *Journal of Comparative Economics* 41, (2), 468–489.

Anderson G.J., I. Crawford and Leicester A. (2011). "Welfare Rankings for Multivariate Data: A Non-Parametric Approach", *Journal of Public Economics*, 95(3-4), 247–252.

Atkinson, A.B., and Brandolini A. (2011), "On the identification of the middle class", ECINEQ WP2011-217.

Banerjee, A.V., and Duflo E. (2008). "What is middle class about the middle classes around the world?", *Journal of Economic Perspectives*, 22(2), 3–28.

Beach, C.M., Richard P. Chaykowski, and Slotsve G. A. (1997). "Inequality and Polarization of Male Earnings in the United States, 1968-1990", *North American Journal of Economics and Finance*, 8(2), 135–151.

Bramall, C. (2001). "The quality of China's household income survey", *China Quarterly*, 167, 689–705.

Brandt, L. and C.A. Holz (2006). "Spatial Price Differences in China: Estimates and Implications", *Economic Development and Cultural Change*, 55(1), 43–86.

Citro, C.F. and Michael, R.T. (1995). *Measuring Poverty: A New Approach*. National Academy Press, Washington, DC.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). "Maximum likelihood from incomplete data via EM algorithm", *Journal of the Royal Statistical Society*, B, 69, 1–38.

Easterly, W. (2001). "The Middle Class Consensus and Economic Development"", *Journal of Economic Growth*, 6(4), 317–35,

Fang, C., E. Wailes and Cramer G. (1998). "China's Rural and Urban Household data: Collection. Availability, and problems", CARD Working Paper 98-WP 202.

Farcomeni, A. (2012) "Quantile Regression for longitudinal data based on latent Markov subject-specific parameters", *Statistics and Computing*, 22, 141–152.

Fitoussi J-P, A. Sen and J.E. Stiglitz (2011). "Mismeasuring Our Lives: Why GDP Doesn't Add Up", The New Press.

Foster, J.E. (1998). "Absolute versus Relative Poverty", *American Economic Review*, Vol. 88(2), Papers and Proceedings 335–341.

Foster, J.E., Greer, J. and Thorbecke, E. (1984). "A Class of Decomposable Poverty Measures ", *Econometrica*, 42, 761–766.

Gibrat, R., (1930), "Une Loi Des Repartitions Economiques: L'effet Proportionelle", *Bulletin de Statistique General*, France, 19, 469.

Gibrat, R., (1931). *Les Inegalites Economiques.* Libraire du Recueil Sirey, Paris.

Goldthorpe, J.H. (2010). "Analysing Social Inequality: A Critique of Two Recent Contributions from Economics and Epidemiology", *European Sociological Review*, 26(6), 731–744.

Gong, C.H. and X. Meng (2008). "Regional Price Differences in Urban China 1986-2001: Estimation and Implication", IZA Discussion Paper n.3621.

Grusky, D. and Kanbur, R. (eds.) (2006). *Poverty and Inequality*, Stanford University Press, Stanford, California.

Hamilton, B.W. (2001). "Using Engel's Law to Estimate CPI Bias", *American Economic Review*, 91, 3, 619-630.

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

Kakwani, N. and J. Silber (2008). *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave Macmillan.

Keribin, C. (2000). "Consistent estimation of the order of mixture model", *Sankya, Series A, Indian Journal of Statistics*, 62, 49–66.

Lo, Y., Mendell N.R. and Rubin D.B.(2001). "Testing the number of components in a normal mixture", *Biometrika*, 88(3), 767-778.

Longford, N. T. (2008). *Studying Human Populations.* Springer, New York.

Marron, J.S. and Wand M.P. (1992). "Exact mean integrated squared error", *The Annals of Statistics*, 20, 712–736.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley, New York.

Melnykov, V. and Maitra, R. (2010). "Finite Mixture Models and Model-Based Clustering", *Statistics Surveys*, 4, 80–116.

Nussbaum, Martha C. (2011). *Creating Capabilities: The Human Development Approach*, Harvard University Press.

Oakes, D. (1999). "Direct calculation of the information matrix via the EM algorithm", *Journal of the Royal Statistical Society, Series B*, 61, 479–482.

Pittau, M.G., Zelli, R. and Johnson P.A. (2010). "Mixture Models, Convergence Clubs and Polarization", *Review of Income and Wealth*, 56(1), 102–122.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ravallion, M. and Chen S. (2007). "China's (uneven) progress against poverty", *Journal of Development Economics*, 82(1), 1–42.

Saez E. and Michael R. Veall (2005). "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence", *The American Economic Review*, 95(3), 831-849.

Sen, A.K. (1976). "Poverty: an ordinal approach to measurement", *Econometrica*, 44(2), 219–231.

Sen, A.K. (1983). "Poor Relatively Speaking", *Oxford Economic Papers*. 35, 153–169.

Sutton, J. (1997). "Gibrat's Legacy", *Journal of Economic Literature* 35, 40–59.

Tao Yang, D. (1999). "Urban-Biased Policies and Rising Income Inequality in China", *American Economic Review*, 89,2, Papers and Proceedings, 306–310.

Townsend, (1985). "A Sociological Approach to the Measurement of Poverty - A Rejoinder to Professor Amartya Sen", *Oxford Economic Papers*, 37, 659–668.

Wu, X. and J.M. Perloff (2005). "China's Income Distribution, 1985-2001", *Review of Economic and Statistics*, 87, 763-775.

Yitzhaki, S. (1994). "Economic Distance and overlapping of distributions", *Journal of Econometrics*, 61 147–159.

Appendix A.1: The urbanization index

The urbanization index is calculated as the proportion of the population living in urban areas normalized at 1990. Therefore, a value greater than 1 shows that the rate of urbanization in that province is greater than the average rate of urbanization across the six provinces in 1990. Urban populations for the six provinces were available for the years 1990, 1994 and 1999. To interpolate and extrapolate indices of urbanization over the observation period quadratics in time were fitted for each province over those years. The resulting indices for the six provinces are presented in Table A.1.

Table A.1: *Urbanization index in the six provinces – base year 1990*

Year	Shaanxi	Jilin	Shangdong	Hubei	Guangdong	Sichuan
1990	0.3966	0.7324	1.7918	1.2884	0.7801	1.0104
1991	0.4451	0.7557	1.9739	1.3550	1.1917	1.2473
1992	0.4872	0.7769	2.1311	1.4198	1.5460	1.4553
1993	0.5228	0.7961	2.2635	1.4830	1.8430	1.6345
1994	0.5520	0.8132	2.3711	1.5446	2.0826	1.7848
1995	0.5747	0.8282	2.4539	1.6045	2.2650	1.9063
1996	0.5909	0.8412	2.5118	1.6628	2.3901	1.9989
1997	0.6007	0.8521	2.5448	1.7194	2.4579	2.0626
1998	0.6040	0.8609	2.5531	1.7743	2.4684	2.0975
1999	0.6008	0.8676	2.5364	1.8276	2.4216	2.1035
2000	0.5911	0.8723	2.4950	1.8793	2.3175	2.0806
2001	0.5750	0.8749	2.4287	1.9293	2.1561	2.0289

Appendix A.2: Expected information matrix through Oakes identity

Let for ease of notation

$$\tilde{\ell}_c = \mathbb{E}_{\mathbf{\Psi}^{(\nu)}} \left\{ \ell_c \left(\mathbf{\Psi} | \mathbf{X}, C \right) \right\}.$$

Oakes (1999) obtains the expected information matrix $\mathbf{J}(\mathbf{\Psi})$ as

$$\mathbf{J}(\mathbf{\Psi}) = -\frac{\partial \tilde{\ell}_c}{\partial^2 \mathbf{\Psi}} - \frac{\partial \tilde{\ell}_c}{\partial \mathbf{\Psi} \partial \mathbf{\Psi}^{(\nu)}},$$

evaluated at the MLE.

After some algebra, it can be seen that the first summand is a block diagonal matrix, with zeros for derivatives regarding elements of different components of the mixture, and where the vj-th element is given by

$$z_{iv}\frac{w_{ij}z_{ij}\sum_{h\neq j}e^{\mathbf{z}_i\boldsymbol{\beta}_h}}{\sum_h e^{\mathbf{z}_i\boldsymbol{\beta}_h}}$$

when v and j identify two elements of the $\boldsymbol{\beta}$ vector belonging to the same component of the mixture.

Let now for ease of notation

$$\phi_{ij} = f\left(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right).$$

The second summand is also given by a block-diagonal matrix, with

$$z_{iv}\frac{w_{ij}z_{ij}\phi_{ij}\sum_{h\neq j}\phi_{ih}e^{\mathbf{z}_i\boldsymbol{\beta}_h}}{\sum_h \phi_{ih}e^{\mathbf{z}_i\boldsymbol{\beta}_h}}$$

whenever v and j identify two elements of the $\boldsymbol{\beta}$ vector belonging to the same component of the mixture.