The progress and characteristics of poor, middle and rich classes in urban China: Results from partial definition of class membership.

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

July 7, 2012

Abstract

Semi-parametric estimation of the components of a mixture distribution describing household income is employed to estimate the probability that a household belongs to a particular class. Various household characteristics are then used to model the determinants of class membership. The methodology is employed to examine the progress of poor, middle and rich classes in Urban China in the last decade of the 20th century.

Keywords: Poverty Frontiers, Mixture Models, Middle Class, Partial Definition.
JEL classification: C14; I32; O1.

[*]Department of Economics, University of Toronto, email: anderson@chass.utoronto.ca
[†]Department of Statistical Sciences, Sapienza University of Rome,
email: grazia.pittau@uniroma1.it, roberto.zelli@uniroma1.it.

1 Introduction.

Leading economic journal paper titles like, "Gaining Ground: Poverty in the Postwar United States", "What is middle class about the middle classes around the world?", "Empirics for growth and distribution: Stratification, polarization, and convergence clubs.", "The world distribution of income: Falling poverty and ... convergence, period." are testament to the established practice in the economics profession of classifying agents within a society into groups in order to study their wellbeing or behavior. Invariably this has involved specifying boundaries or frontiers for set 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. Here a technique is proposed for categorizing agents without resort to such boundaries, rather an agent's category is partially determined by the behavior they exhibit with respect to an economic variable. 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.

With regard to disputed boundary formation, historically the poor have probably received most attention, large quantities of ink have been spilled over how we may identify and study the poor. Following Sen's seminal paper (Sen, 1976) the income poor are 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 whereas agents above the cut-off are classified as non-poor. There has been a large and extensive literature on how the cutoff should be determined with the Sen (1983) versus Townsend (1985) relative versus absolute debate being a feature (Foster, 1998). Townsend advocated a relative measure, usually a proportionate to median income line, for example 50% of median income, Sen advocated an absolute measure, the U.N. $1 and $2 a day measures are an example, Citro and Michael (1995) proposed a combination of the two. Generally absolute measures reflect a "needs" based view of the poverty frontier and are frequently based upon calculation of the budget needed to satisfy a set of basic needs. Relative measures on the other hand appear to be based upon a public view of what is socially acceptable (Hills, 2002), (This view is not recent, Adam Smith is often interpreted as having a "relative" sense of necessity when he wrote: "$\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)). Whichever approach is taken it will have a quality of arbitrariness about it.

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 incomes. For example Easterly (2001) defines the "middle class" as those lying between the 20th and 80th percentile on the consumption distribution. 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 below a scaled down value of median wages using 0.25, 0.5 and 0.75 of the median as frontiers and the proportions a people above 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 or identification technique has taken the form of arguing that "poorness" and indeed "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. Other strands of literature see class as an inherited or tribal trait, a matter of how one speaks or eats. Indeed this lead Atkinson and Brandolini argued 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."

What is clear is that 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. So it is quite possible for two agents with the same income, or the same job, or the same wealth to come from ostensibly different classes. The problem with expanding the domain of characteristics for determining class membership is that determining class frontiers over many dimensions compounds the difficulties with definition and intensifies 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 their observable actions.

There is an extensive theoretical literature on the size distribution amongst agents characterized by a variable x that is the consequence of a stochastic process of the form:

$$x_t = x_{t-1} + \psi + e_t.$$

Here the process is assumed to have started at $t = 0$ and had run for T periods where e_t was assumed to be a serially independent zero mean and finite variance shock. Very often the size distribution turns out to be 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; Sutton, 1997). 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 shocks though the mean and variance of the distribution do depend on the parameters governing the process. Suppose that the functionings and capabilities set that characterize a particular class (denote it "i") also determines the parameters that govern the stochastic process of an observable variable 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 the variable x, x_t will have a particular distribution $f_i(x_t)$ that is distinguishable from the distribution of $f_j(x_t)$ for class $j \neq i$. Furthermore the distribution of 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 $P_i(x_t)$, the probability that an agent with income x_t in period t is a member of the i'th class $i = 1, \cdots, K$, in that period. Thus class membership is only partially identified in a sense similar to Manski (2003) in respect of probability distributions.

Here following Banerjee and Duflo (2008) these ideas are exploited to identify and examine trends in economic class structure in urban China in the last decade of the 20th century and to examine what is typical about class membership. Banerjee and Duflo (2008) were interested in what is typical about the middle class, they examine a number of household characteristics over samples from several developing countries to see if there are any distinguishing features. They consider expenditure patterns for example, what do middle class households spend on eating, drinking, entertainment, education, healthcare and housing. They look at occupational patterns, what sort of jobs they have, are they typically professional or entrepreneurial for example. They examine the finances of the households (are they credit constrained?) and whether or not they are geographically mobile. Finally they consider the family size, fertility, education and savings choices of households. They conclude that middle class households have a "good job", they have fewer children and spend more on healthcare and education of their children and their own healthcare all of which are facilitated by having a good job.

In addition to these factors 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).

The remainder of the paper is organized as follows. In the next section, we briefly describe our approach in which class membership can only be partially identified (i.e. only the probability of a household being in a class can be estimated). Section 3 reports the main empirical results. Section 3.1 presents a brief description of the sample drawn from the Urban Household Survey, spanning the period from 1992 to 2001. In section 3.2 representation, interpretation and estimation of mixture models are illustrated. It appears that four classes corresponding to the poor, lower middle upper middle and rich income classes emerge throughout the period. Class sizes, income growth and the extent of class identification are examined. In section 3.3 the association between class membership probability and various characteristics of the household is studied. Section 4 summarizes and concludes.

2 Partial definition of group membership.

If it is assumed that there are a finite number of classes in society whose behaviors are governed by their circumstances to the extent that their income paths follow distinct processes, then the income size distribution of the i'th group at time t, $f_i(x_t)$ will be distinct from the j'th groups distribution $f_j(x_t)$. With K such groups in a society the overall income size distribution $f(x_t)$ will be a mixture of these where the mixing weights w_i correspond to the population shares of the respective classes so that:

$$f(x_t) = \sum_{i=1}^{K} w_i f_i(x_t); \text{ where } \sum_{i=1}^{K} w_i = 1. \tag{1}$$

Given some assumptions regarding the nature of the $f_i(x_t)$'s (normality or log normality are popular specifications that can be theoretically rationalized for example), their parameters and the values of the class shares can be estimated. These in turn facilitate estimates of the probability that an agent with an income x^* in period t is in the i'th group in that period since:

$$P_i\left(x_t^*\right) = P\left(x_t^* \in i'\text{th class}\right) = \lim_{\Delta\to 0}\left(\frac{P\left(x \in x_t^*,\, x_t^* + \Delta \,\text{for}\, i'\text{th class}\right)}{P\left(x \in x_t^*,\, x_t^* + \Delta \,\text{for population}\right)}\right) =$$

$$= \lim_{\Delta\to 0}\left(\frac{\int_{x_t^*}^{x_t^*+\Delta} w_i f_i(x)\,dx}{\int_{x_t^*}^{x_t^*+\Delta}\sum_{i=1}^{K} w_i f_i(x)\,dx}\right) = \frac{w_i f_i(x_t^*)}{\sum_{i=1}^{K} w_i f_i(x_t^*)},\ \text{for}\, i = 1\cdots K. \tag{2}$$

Effectively this provides K group membership indices for each agent in the population.

How well determined are the respective groups? Note that it is possible for the group distributions to overlap, that is for an agent with income x^* 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 income 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 income will only partially identify her group membership in the sense that the probability of her being in a particular group is all that can be obtained. If for example one of the group distributions (suppose it to be the j'th) was not overlapped by any other group distribution, that is all other group distributions had no support in the domain of the jth distribution, then j'th group membership would be defined with probability one since:

$$P\left(x_t^* \in j'\text{th class}\right) = \frac{w_j f_j(x_t^*)}{\sum_{i=1}^{K} w_i f_i(x_t^*)} = 1,\ \text{for}\, x^* \in f_j\ \text{support}, (\text{since}\, f_i(x^*) = 0,\ \forall i \neq j)$$

When there is overlap, an index of the extent of the identification (II) of the j'th group can be generated by calculating the complement to one of the area of overlap with the other (relatively weighted) distributions as follows:

$$II_j = 1 - \int_{-\infty}^{+\infty} \min\left(f_j(x),\, \frac{1}{w_j}\sum_{\substack{i=1\\ i\neq j}}^{K} w_i f_i(x)\right)\, dx. \tag{3}$$

For example letting x^* be the intersection point of the two weighted normals $\mathrm{N}\left(\mu_p, \sigma_p^2\right)$ and $\mathrm{N}\left(\mu_r, \sigma_r^2\right)$ (rich and poor groups) with respective weights w and $1-w$

where $\mu_p < x^* < \mu_r$ then the overlap measure in terms of the poor distribution is given by:

$$OV = (1 - \Phi((x^* - \mu_p)/\sigma_p)) + (1 - w)\Phi((x^* - \mu_r)/\sigma_r)/w, \tag{4}$$

where Φ the cumulative normal distribution function and x^* the solution to

$$w \frac{\mathrm{e}^{\frac{-(x^*-\mu_p)^2}{2\sigma_p^2}}}{\sqrt{2\pi\sigma_p^2}} = (1 - w) \frac{\mathrm{e}^{\frac{-(x^*-\mu_r)^2}{2\sigma_r^2}}}{\sqrt{2\pi\sigma_r^2}}$$

which may be written as the following quadratic form in x^* with the root between μ_p and μ_r providing the intersection point:

$$\left(\frac{1}{\sigma_r^2} - \frac{1}{\sigma_p^2}\right) x^{*2} - 2\left(\frac{\mu_r}{\sigma_r^2} - \frac{\mu_p}{\sigma_p^2}\right) x^* + \left(\frac{\mu_r^2}{\sigma_r^2} - \frac{\mu_p^2}{\sigma_p^2}\right) - 2\ln\left(\frac{(1-w)\sigma_p}{w\sigma_r}\right) = 0. \tag{5}$$

What are other characteristics of agents in a particular group? The probability measures can serve as selectors which permit the calculation of a whole range of class characteristics. Suppose an agent with income x_t at time t reports the status of another characteristic z (suppose for example it is health or wealth index) as z_t, then a whole range of indices for the status of the i'th class with respect to z can be calculated for example means, variances and Foster-Greer-Thorbecke (FGT, 1984) generalized measure of poverty with respect to a cutoff z^* would be:

$$\overline{Z_i} = \frac{1}{n\,w_i} \sum_{t=1}^{n} P_i(x_t)\, z_t$$

$$\mathrm{V}(Z_i) = \frac{1}{n\,w_i} \sum P_i(x_t) \left(z_t - \overline{Z_i}\right)^2$$

$$\mathrm{FGT_M}(Z_i) = \frac{1}{n\,w_i} \sum_{t=1}^{n} P_i(x_t) \mathrm{I}\left(z^* - z_t\right) \left(\frac{z^* - z_t}{z^*}\right)^{\mathrm{M}-1};$$

where

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

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 i and group j can be made by considering the incomplete subgroup moments:

$$\mathrm{F}_i(z^*) - \mathrm{F}_j(z^*) = \frac{1}{n}\sum_{t=1}^{n}\left[\left(\frac{P_i(x_t)}{w_i} - \frac{P_j(x_t)}{w_j}\right)(z^* - z_t)^{\mathrm{M}-1}\,\mathrm{I}\left(z_t \leq z^*\right)\right].$$

What characteristics determine group membership? The P_i's can themselves be the object of description. Suppose z_t is a vector of circumstances of agent with income x_t at time t, estimates of β_i in relationships of the form:

$$P_i(x_t) = g_i\left(\beta_i, z_t\right) + e_i, \;\; i = 1 \cdots K \tag{6}$$

provide information on the determinants of class membership. Note that in this system of equations the dependent variables sum to one and reside in the unit interval so that the $g_i(.)$'s and e_i's will have to satisfy the appropriate constraints much like systems of demand equations which describe expenditure shares[1].

3 Evolution of the middle class 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 (the average annual growth rate of city incomes over the period 1990-1999 was over 18%). In addition there was a massive migration to the cities (in 1985 20% of the Chinese population was urbanized, by 1999 over 42.6% 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.

[1]Note that the regression coefficients obtained from regressing the P's on the z's has the simple interpretation of reweighted class averages of the z's where the reweighting matrix is the covariance matrix of the z's.

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[2].

Among the six provinces, two are ranked as poor economies, two as upper-middle and two rich economies. Specifically, the six selected provinces are characterized by different levels of per capita GDP: Shaanxi with per capita GDP in the year 2010 equal to 3,966 US $ (source: China's statistics yearbook), Sichuan (per capita GDP equal to 3,104 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 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

[2] 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.

Gong and Meng (2008) SPI as regional deflators. Gong and Meng have recently 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[3].

3.2 Modelling income distribution.

In each year, income data are interpreted as a sample from a mixture of k components in unknown proportions $\pi_1, \cdots, \pi_k$, that are nonnegative and sum to one. The probability density function of the random vector X_t under a k-component mixture model is defined as:

$$f(x_t, \mathbf{\Psi}) = \sum_{i=1}^{k} \pi_i f_i(x_t, \theta_i), \tag{7}$$

where the vector $\mathbf{\Psi} = (\pi_1, \cdots, \pi_{k-1}, \xi')'$ contains all the unknown parameters of the mixture model; π_i, $i = 1, \cdots, k$ represent the mixing proportions and the vector ξ contains all the parameters $(\theta_1, \cdots, \theta_k)$ known *a priori* to be distinct; $f_i(x_t, \theta_i)$ denotes the values of the univariate density specified by the parameter vector θ_i. Each component represents the income distribution of a homogeneous group of households, that is a household belonging to group i faces income opportunities described by the distribution f_i. The mixing proportion π_i $(i = 1, \cdots, k)$ gives the prior probability that an economic unit belongs to the ith component of the mixture. It is an endogenous parameter which determines the relative importance of each component in the mixture.

One of the main advantage in using mixture models is that, once a model is generated, posterior probabilities that a household with income x_t comes from a component of the mixture can be computed. Formally, the posterior or conditional probability τ_{it} is:

$$\tau_{it} = \tau_i(x_t, \mathbf{\Psi}) = \frac{\pi_i f_i(x_t, \theta_i)}{\sum_{i=1}^{k} \pi_i f_i(x_t, \theta_i)}, \tag{8}$$

which represents the probability that the household with income x_t belongs to the i-th component of the mixture.

[3]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).

We took the component densities to be normal[4], with the number of components to be established. Therefore, each component is characterized by its mean and variance: $\theta_i = (\mu_i, \sigma_i^2)$.

The unknown parameters (means, variances and proportions of each component) are estimated by maximum likelihood (ML) via the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977). Starting from a given number of components and an initial parameter $\boldsymbol{\Psi}^{(0)}$, the first stage of the algorithm (E-step) is to assign to each data point its current posterior probabilities. In the second stage (M-step), the maximum likelihood estimates are computed using the posterior probabilities as conditional mixing weights. The estimates of the parameters are used to re-attribute a set of improved probabilities of group membership and the sequence of alternate E and M steps continues until a satisfactory degree of convergence occurs to the ML estimates[5].

In modelling households income distribution by using a finite mixture model, special attention should be paid to the choice of the appropriate number of the components of the mixture that represent distinctive sub-populations from which the sample arises. To assess the number of components, we conducted a likelihood ratio (LR) test of the null hypothesis that a random sample is from a k_0-component mixture ($H_0 : k = k_0$) versus the alternative $H_1 : k_1 = k$ for some $k_1 > k_0$. Since regularity conditions do not hold, the LR statistic does not have an asymptotic chi-squared null reference distribution. Lo *et al.*(2001) showed that the LR statistic is asymptotically converging to a weighted sum of $p+q$ independent chi-squared with one degree of freedom, where $p = 3k_1 - 1$ and $q = 3k_0 - 1$. Since the convergence rate of 2LR to the limiting distribution is slow, Lo *et al.* (2001) suggest using an adjusted likelihood ratio statistic

[4]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 twofold 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. The assumption of normality, in fact, results from additive shocks to the expected income of each strata.

[5]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. In this paper, randomly selected starts, large sample non-hierarchical (Kaufman and Rousseeuw, 1990) clustering-based starts have been selected for initialization.

(with factor of adjustment $(1+1/[(p-q)log(n)])$) to achieve reasonable accuracy. Lo (2005) shows that this test works reasonably well for testing the number of components in a heteroscedastic normal mixture [6].

We apply the LR test sequentially beginning with the null hypothesis $k_0 = 1$ against the alternative $k_1 = 2$, continuing to that of $k_0 = 2$ against the alternative $k_1 = 3$, and so on. Table 1 presents the results of the testing procedure, starting with the null $k_0 = 2$, since the null hypothesis of one component is always strongly rejected. The two-component model is always rejected in favor of the model with at least three components. A three-component mixture seems to be the 'best' parsimonious model at the beginning of the period (1992 and 1993) and also in 1998. For the remaining years a four-component mixture is the most suitable model. At any rate, adding a fifth component to the mixture does not improve the fitting of the model. Although at the beginning of the period the components are not separated enough to give rise to the same number of components detected in the subsequent years, to ease the interpretation we always adopted the four-component solution. Our choice of four-component mixture is corroborated by the fact that each component is always characterized by distinct means, relatively modest dispersion and not negligible size. No bizarre situations appear in the model fits, like the presence of clusters with very small variance, very flat components with large dispersion and very small marginal probabilities, components with similar means but different shape due to their disparate variances, etc., i.e. components that can play a role in improving the fit of the whole distribution but may be unacceptable in terms of economic interpretability. 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. As the mixing proportions (π) indicate, none of the groups is dominant, that is no component persistently accounts for more than half of the households.

Figure 1 show the pattern of the log-likelihhod (with inverted sign) of the estimated mixture model from one to five components relative to the year 2001. As evident the

[6] Another approach, fitting-oriented, is to find a criterion that assess the number of components according to the mixture that 'best' fits the data. See Pittau *et al.* (2010) for the implementation of a test based on the goodness of fit of the estimated mixture model by comparison of a kernel estimate of the density of the data and its expected value under the null hypothesis that the population density is a mixture.

Table 1: *The choice of the number of components according to the adjusted likelihood ratio test.*

	k_0=2 vs k_1=3		k_0=3 vs k_1=4		k_0=4 vs k_1=5	
year	LR*	p-value	LR*	p-value	LR*	p-value
1992	54.5	0.00	10.3	0.49	8.7	0.99
1993	109.8	0.00	17.3	0.14	20.4	0.53
1994	141.6	0.00	23.1	0.02	21.2	0.44
1995	89.7	0.00	35.8	0.00	4.3	0.99
1996	73.8	0.00	21.4	0.06	7.3	0.99
1997	83.6	0.00	18.8	0.09	3.7	0.99
1998	69.9	0.00	11.3	0.38	0.8	1.00
1999	74.9	0.00	19.9	0.08	7.4	0.85
2000	84.7	0.00	38.8	0.00	18.6	0.64
2001	87.4	0.00	67.5	0.00	14.4	0.79

improvement is large moving from one to three components and it is also sizeable passing from three to four, while it is negligible adding a fifth component.

Table 2 provides a summary of the model fits for all years: the estimated mean (μ) and standard deviation (σ) of each normal component along with its corresponding mixing proportion (π). Figure 2 visually compares fitted three component mixtures for all the years of the analysis with the corresponding estimated kernel density[7]

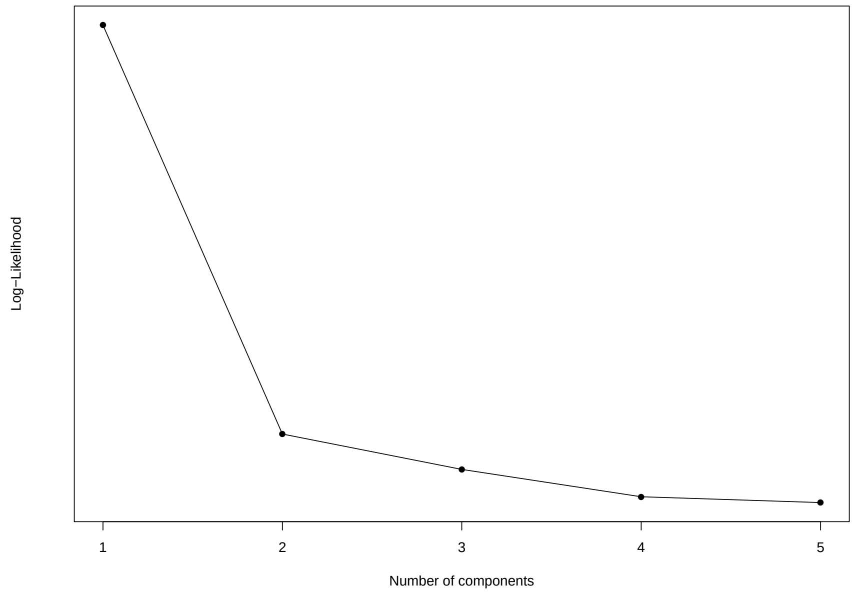
Table 2: *Estimated parameters of the components of the mixtures*

	Poor			Lower-middle			Upper-middle			Rich		
Year	μ	σ	π	μ	σ	π	μ	σ	π	μ	σ	π
1992	917	238	0.13	1400	281	0.27	1914	503	0.49	3348	524	0.11
1993	921	241	0.10	1465	395	0.32	2260	625	0.47	3980	707	0.11
1994	753	177	0.05	1369	367	0.30	2356	668	0.53	4255	771	0.12
1995	921	241	0.10	1465	395	0.32	2260	625	0.47	3980	707	0.11
1996	1087	280	0.10	1779	450	0.36	2644	683	0.42	4421	789	0.11
1997	1036	271	0.09	1682	420	0.29	2556	690	0.49	4340	797	0.12
1998	1196	352	0.10	1936	583	0.35	2910	971	0.48	5421	745	0.07
1999	1318	401	0.14	2208	610	0.40	3432	954	0.38	5823	852	0.09
2000	1295	420	0.15	2316	656	0.43	3786	1021	0.34	6451	1049	0.09
2001	1367	432	0.15	2553	752	0.50	4504	1307	0.31	8058	898	0.04

Note: μ and σ are expressed in 1994 constant yuan. π are the mixing proportions.

[7]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 h is the estimated bandwidth of the kernel.

Figure 1: *Log-likelihood (–) of the estimated mixture model from one to five components. Year 2001*

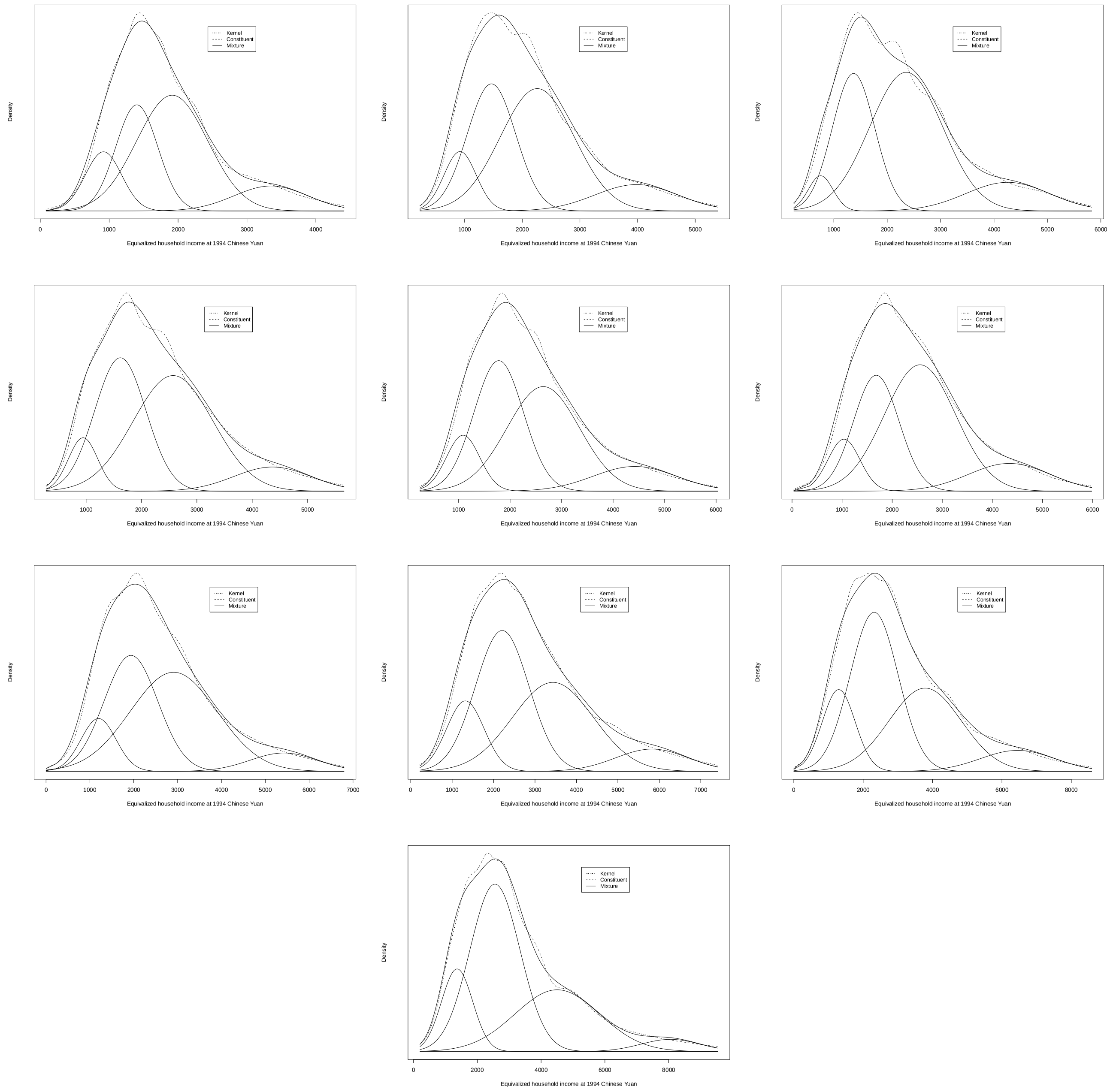


Real income annual growth rates of 4.54% for the poor, 6.90% and 9.98% for the lower and upper middle class respectively, and 10.25% for the rich. The probability of being in the poor group are stable over the period (from 12.9% to 15.2%), while the lower-middle income group probability grows (from 26.7% to 49.6%) and the upper-middle income group probability diminishes (from 49.3% to 30.9%). The rich group mixing proportions falls from 11.2% to 4.3%. The within group inequalities grow for all groups (which is consistent with Gibrat's law).

As previously described, in the mixture approach the membership of each class is not determined with certainty, but each household has attached an estimated probability of belonging to each component of the mixture. These probabilities are function of the location of each household's income value. As the components overlap, there is considerable uncertainty about the household's allocation, while if the components are well separated, the conditional probabilities τ_{it} tend to define a partition/segmentation of the population.

Based on equation (3), we can contemplate the index of determination (i.e. the degree of distinction) of each group. The following table reports the results for each

Figure 2: *Kernel density estimation and the (inflated) four-components mixture model fit, 1992–2001.*



year. Upper middle and Rich groups appear to be more differentiated (less overlapped) and thus better determined than the poor and lower middle classes and there is an upward trend in the determination of all classes over the sample period consistent with a steady increase in the polarization between all groups so that the four groups are well distinguished by the end of the period.

Table 3: *Identification index of each group*

Year	Poor	Lower-Middle	Upper-Middle	Rich
1992	49.1	3.4	53.4	93.1
1993	24.6	22.7	55.3	91.8
1994	24.3	40.3	65.8	92.4
1995	24.6	22.7	55.3	91.8
1996	30.0	33.0	49.3	90.0
1997	33.2	21.3	56.1	90.4
1998	14.3	13.1	49.0	94.4
1999	32.4	36.1	49.0	91.3
2000	38.1	47.2	51.2	90.7
2001	36.3	57.8	58.7	95.9

As described in Section 3, it is possible to calculate a whole range of indices for the status of the income classes with respect to a list of characteristics of the households. For example, Tables 4, 5, 6 report the means of household size, age of the household head, and urbanization index for each class over time. From the household size table we see that generally poor and lower middle class status families tend to be larger though household size is declining for all classes. The age of household head makes a clear distinction between the upper and lower middle classes, typically upper middle class households have older heads than lower middle class families. The urbanization index table clearly indicates that the poorer classes are more prevalent in provinces with lower urbanization rates.

3.3 Modelling the conditional probabilities.

Estimating a system of equations that relate to the posterior probabilities τ_{it} may help identify which factors significantly contribute in explaining the degree of association of each household to the components. The dependent variables (i.e. the probabilities τ_{it}) are naturally restricted to lie between zero and one and their sum with respect to i is

Table 4: *Households size in the income classes*

Year	Poor	Lower-Middle	Upper-Middle	Rich
1992	3.963	3.486	2.600	3.229
1993	3.822	3.468	2.655	3.162
1994	3.955	3.472	2.712	3.146
1995	3.834	3.384	2.680	3.072
1996	3.790	3.338	2.687	3.083
1997	3.722	3.352	2.786	3.102
1998	3.595	3.288	2.693	3.049
1999	3.538	3.218	2.640	2.987
2000	3.594	3.218	2.622	2.941
2001	3.558	3.183	2.574	2.870

Table 5: *Age of the household head in the income classes*

Year	Poor	Lower-Middle	Upper-Middle	Rich
1992	43.661	42.708	50.072	43.980
1993	44.037	43.531	50.346	44.632
1994	45.717	44.119	48.523	44.101
1995	45.959	44.128	48.714	44.876
1996	45.315	43.872	47.938	44.775
1997	44.344	43.624	47.288	44.426
1998	44.530	44.213	47.035	44.904
1999	44.832	44.322	47.848	45.389
2000	45.678	45.287	49.266	46.653
2001	46.147	45.564	49.964	46.954

Table 6: *Urbanization index in the income classes*

Year	Poor	Lower-Middle	Upper-Middle	Rich
1992	1.198	1.339	1.367	1.373
1993	1.223	1.411	1.539	1.535
1994	1.291	1.469	1.663	1.634
1995	1.334	1.576	1.813	1.724
1996	1.567	1.681	1.793	1.753
1997	1.519	1.701	1.847	1.808
1998	1.496	1.728	1.987	1.850
1999	1.616	1.777	1.852	1.854
2000	1.564	1.739	1.894	1.872
2001	1.541	1.723	1.885	1.811

equal to one. Therefore, it is expected that also the modelled probabilities should satisfy these restrictions. These restrictions do not allow one to use a traditional modelling framework based on the multinormal distribution of the stochastic component.

Following Fry *et al.* (1996), to ensure that the stochastic component of the model will satisfy the restriction that the modelled probabilities should be constrained to the unit simplex, the compositional data analysis (hereafter CODA) will be used. The CODA technique relies on the use of "log-ratio" in the statistical analysis (Aitchison, 1986). To surmount the problem that confinement of probabilities to the unit simplex presents, a one-to-one transformation is is employed to map the data on probabilities to transformed data suitable for analysis based on multivariate normal techniques (e.g. multivariate regression). One such transformation is the additive log-ratio defined as:

$$y_{it} = \log\left(\frac{\tau_{it}}{\tau_{kt}}\right), \quad i = 1, \cdots, k-1.$$

The inverse transformation is the additive logistic transform and the probabilities can be obtained as:

$$\tau_{it} = \frac{\exp(y_{it})}{1 + \exp(y_{1t}) + \cdots + \exp(y_{k-1,t})}, \quad i = 1, ..., k-1.$$

$$\tau_{kt} = 1 - \tau_{1t} - \tau_{2t} - \cdots - \tau_{k-1,t}$$

.

It should be noted that the unit sum constraint reduces the dimension of the probability space. However, all the statistical procedures involving the log-ratio covariance matrix are invariant to the probability order and to the choice of the probability used as denominator in the log-ratios (see Aitchison, 1986, p.93-98). To overcome the asymmetry in the treatment of probabilities, the estimating equations have been specified in terms of centered log-ratios by applying the following transformation (McLaren *et al.*, 1995):

$$y_{it} = \log\left(\frac{\tau_{it}}{\tilde{\tau}_t}\right), \quad i = 1, \cdots, k$$

where $\tilde{\tau}_t$ is the geometric mean of the k estimated probabilities associated to the t-th household. After the log-ratio transformation the model can be treated as a multiple linear regression model. The fact that the dependent variable has been estimated in an auxiliary analysis does not necessarily present any difficulties for regression analysis, aside from a possible loss of efficiency. The loss of efficiency is because the additional source of uncertainty due to the measurement error (the difference between the true value of the dependent variable and its estimated value).

Tables 7 reports the results of the CODA for the 1992-2001 sample, where the 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 as: university undergraduates, college, technical secondary school, high school, middle school, primary school (reference group), others), 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[8].

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.

3.3.1 Household location effects

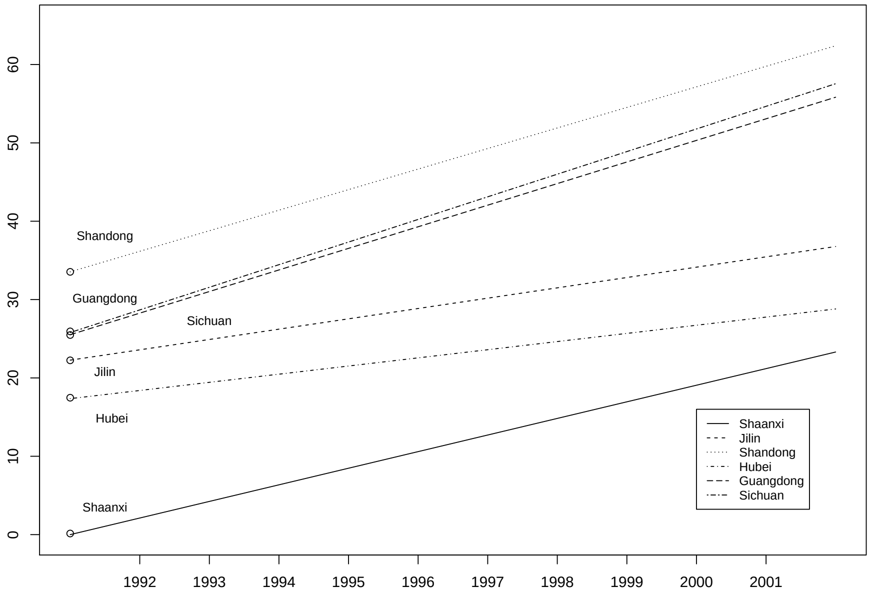
There are strong provincial factors both in levels and trends. Households in all provinces have a higher probability of being poor than the base province (Shaanxi). The provincial trends are for increasing probability of poverty status (relative to the base province) except for Jilin (see Figure 3). The location effects on the probability of belonging to the upper-middle group are similar, but with opposite signs. With respect to provincial urbanization rates the probability of poverty group membership diminishes with rate of urbanization whereas the probability of middle and rich group membership increases. These effects are highly significant and are compounded for pre-OCP families.

[8] See Appendix for details on its construction.

Table 7: *Results of the CODA analysis*

	FIRST COMPONENT					SECOND COMPONENT					THIRD COMPONENT					FOURTH COMPONENT				
	Estimate	s.e.	t val.	P(> |t|)		Estimate	s.e.	t val.	P(> |t|)		Estimate	s.e.	t val.	P(> |t|)		Estimate	s.e.	t val.	P(> |t|	
Shaanxi	-					-					-					-				
Jilin	22.269	1.001	22.240	0.000	***	-2.516	0.172	-14.639	0.000	***	-12.382	0.504	-24.547	0.000	***	-7.371	0.377	-19.531	0.000	***
Shandong	33.539	2.220	15.106	0.000	***	-3.055	0.381	-8.014	0.000	***	-25.004	1.118	-22.355	0.000	***	-5.481	0.837	-6.550	0.000	***
Hubei	17.357	1.369	12.677	0.000	***	-1.707	0.235	-7.260	0.000	***	-12.611	0.690	-18.283	0.000	***	-3.040	0.516	-5.890	0.000	***
Guangdong	25.514	1.812	14.082	0.000	***	-2.593	0.311	-8.335	0.000	***	-18.821	0.913	-20.621	0.000	***	-4.100	0.683	-6.004	0.000	***
Sichuan	25.793	1.501	17.186	0.000	***	-2.796	0.258	-10.853	0.000	***	-17.673	0.756	-23.375	0.000	***	-5.323	0.566	-9.411	0.000	***
time trend	2.119	0.120	17.615	0.000	***	-0.002	0.021	-0.074	0.941		-1.558	0.061	-25.720	0.000	***	-0.559	0.045	-12.327	0.000	***
Shaanxi:trend	-					-					-					-				
Jilin: trend	-0.800	0.154	-5.203	0.000	***	0.176	0.026	6.653	0.000	***	0.223	0.077	2.875	0.004	**	0.402	0.058	6.932	0.000	***
Shandong:trend	0.505	0.160	3.158	0.002	***	-0.095	0.027	-3.459	0.001	***	-0.324	0.080	-4.032	0.000	***	-0.085	0.060	-1.415	0.157	
Hubei:trend	1.078	0.162	6.670	0.000	***	-0.115	0.028	-4.133	0.000	***	-0.770	0.081	-9.459	0.000	***	-0.193	0.061	-3.171	0.002	**
Guangdong:trend	0.637	0.174	3.653	0.000	***	-0.094	0.030	-3.143	0.002	**	-0.520	0.088	-5.919	0.000	***	-0.023	0.066	-0.349	0.727	
Sichuan:trend	0.768	0.156	4.932	0.000	***	-0.049	0.027	-1.825	0.068		-0.671	0.078	-8.547	0.000	***	-0.049	0.059	-0.831	0.406	
Urbanization index	-20.796	1.154	-18.022	0.000	***	2.035	0.198	10.274	0.000	***	15.197	0.581	26.142	0.000	***	3.564	0.435	8.196	0.000	***
age	0.853	0.159	5.371	0.000	***	-0.056	0.027	-2.042	0.041	*	-0.442	0.080	-5.524	0.000	***	-0.355	0.060	-5.936	0.000	***
age2	-0.015	0.002	-8.074	0.000	***	0.001	0.000	3.253	0.001	***	0.008	0.001	8.570	0.000	***	0.006	0.001	8.485	0.000	***
Household size	10.330	0.193	53.413	0.000	***	-0.761	0.033	-22.923	0.000	***	-6.042	0.097	-62.020	0.000	***	-3.526	0.073	-48.379	0.000	***
Single parent dummy	5.851	0.747	7.831	0.000	***	-0.591	0.128	-4.610	0.000	***	-3.290	0.376	-8.741	0.000	***	-1.970	0.282	-6.995	0.000	***
University	-13.657	0.737	-18.527	0.000	***	0.944	0.127	7.457	0.000	***	8.059	0.371	21.701	0.000	***	4.654	0.278	16.753	0.000	***
College	-10.670	0.652	-16.364	0.000	***	0.821	0.112	7.336	0.000	***	6.210	0.328	18.906	0.000	***	3.638	0.246	14.806	0.000	***
Technical secondary school	-7.458	0.646	-11.548	0.000	***	0.558	0.111	5.034	0.000	***	4.376	0.325	13.448	0.000	***	2.524	0.243	10.370	0.000	***
High school	-5.033	0.612	-8.223	0.000	***	0.384	0.105	3.655	0.000	***	2.945	0.308	9.551	0.000	***	1.704	0.231	7.388	0.000	***
Middle school	-2.437	0.596	-4.086	0.000	***	0.200	0.102	1.950	0.051		1.383	0.300	4.604	0.000	***	0.854	0.225	3.800	0.000	***
Primary school	-					-					-					-				
Others	5.470	3.626	1.508	0.131		-0.809	0.623	-1.299	0.194		-2.695	1.827	-1.475	0.140		-1.967	1.367	-1.439	0.150	
SOE employee	-					-					-					-				
Employee of collective enterp.	4.867	0.411	11.838	0.000	***	-0.523	0.071	-7.408	0.000	***	-2.673	0.207	-12.906	0.000	***	-1.671	0.155	-10.785	0.000	***
Other Employee, Self Employed	-0.490	0.590	-0.831	0.406		0.323	0.101	3.189	0.001	***	-0.348	0.297	-1.171	0.242		0.515	0.222	2.316	0.021	*
Retired	1.931	0.648	2.978	0.003	***	-0.089	0.111	-0.797	0.425		-1.130	0.327	-3.459	0.001	***	-0.712	0.244	-2.915	0.004	**
House workers, unemp., students	4.622	1.366	3.384	0.001	***	0.176	0.234	0.752	0.452		-3.743	0.688	-5.440	0.000	***	-1.056	0.515	-2.051	0.040	*
Employed after retirement	-4.573	1.454	-3.145	0.002	***	0.143	0.250	0.573	0.566		3.111	0.732	4.248	0.000	***	1.318	0.548	2.406	0.016	
Pre OCP dummy variable (dva)	48.346	30.730	1.573	0.116		-12.805	5.276	-2.427	0.015	**	-21.955	15.481	-1.418	0.156		-13.586	11.582	-1.173	0.241	
Urbanization index:dvar	-1.136	0.495	-2.294	0.022	*	0.438	0.085	5.156	0.000	***	0.176	0.249	0.704	0.481		0.522	0.187	2.797	0.005	**
hhsize:dvar	1.986	0.323	6.150	0.000	***	-0.192	0.055	-3.456	0.001	***	-0.885	0.163	-5.443	0.000	***	-0.909	0.122	-7.469	0.000	***
Single parent dummy:dvar	-4.391	1.338	-3.282	0.001	***	0.598	0.230	2.606	0.009	**	2.436	0.674	3.614	0.000	***	1.357	0.504	2.691	0.007	**
University:dvar	-					-					-					-				
College:dvar	-2.196	1.338	-1.641	0.101		0.330	0.230	1.434	0.152		0.782	0.674	1.160	0.246		1.085	0.504	2.150	0.032	*
Technical secondary school:dvar	-2.026	1.413	-1.434	0.152		0.191	0.243	0.787	0.431		0.979	0.712	1.374	0.169		0.857	0.533	1.609	0.108	
High school:dvar	-1.679	1.218	-1.379	0.168		0.156	0.209	0.748	0.455		0.790	0.614	1.288	0.198		0.733	0.459	1.596	0.111	
Middle school:dvar	-3.973	1.131	-3.513	0.000	***	0.318	0.194	1.637	0.102		2.043	0.570	3.586	0.000	***	1.612	0.426	3.781	0.000	***
Primary school:dvar	-1.736	0.993	-1.749	0.080		0.102	0.170	0.597	0.551		0.987	0.500	1.974	0.048	*	0.647	0.374	1.730	0.084	
Others:dvar	-2.353	3.998	-0.589	0.556		0.380	0.686	0.553	0.580		1.036	2.014	0.515	0.607		0.937	1.507	0.622	0.534	
SOE employee:dvar	-					-					-					-				
Employee of collective enterp.:dvar	1.099	2.639	0.416	0.677		-0.075	0.453	-0.165	0.869		-0.399	1.330	-0.300	0.764		-0.625	0.995	-0.628	0.530	
Other employee, Self Employed:dvar	8.623	3.939	2.189	0.029	*	-0.530	0.676	-0.784	0.433		-4.462	1.984	-2.248	0.025	*	-3.631	1.485	-2.446	0.014	**
Retired:dvar	-2.996	1.181	-2.537	0.011	*	0.560	0.203	2.763	0.006	**	1.438	0.595	2.417	0.016	*	0.998	0.445	2.243	0.025	**
House workers, unemp, students:dvar	3.241	2.693	1.204	0.229		-0.403	0.462	-0.872	0.383		-1.344	1.357	-0.991	0.322		-1.494	1.015	-1.472	0.141	
Employed after retirement:dvar	-1.422	1.938	-0.734	0.463		0.397	0.333	1.194	0.232		0.188	0.976	0.193	0.847		0.836	0.731	1.145	0.252	

Figure 3: *CODA coefficients of the provinces and their trend - First component*



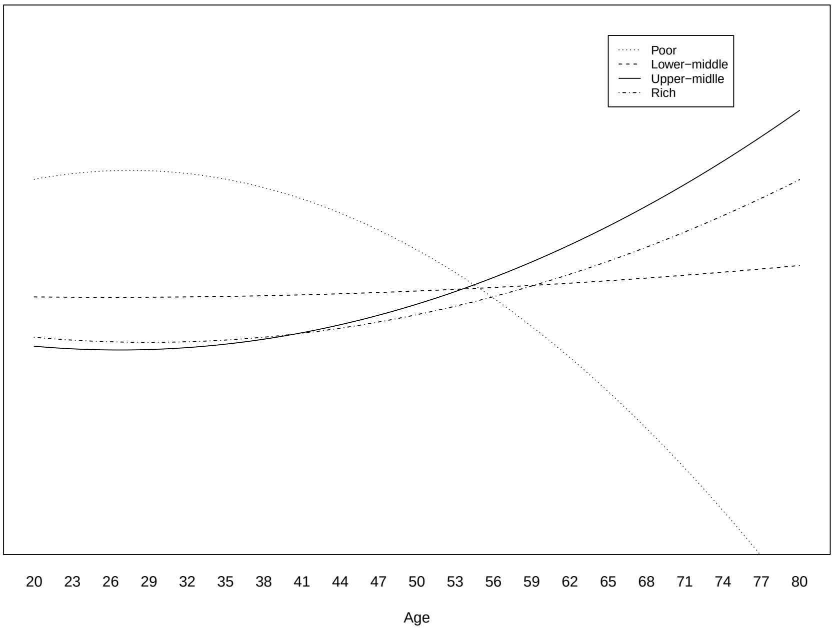
3.3.2 Household characteristic effects

The maturity or age of the household is measured by the age of the household head. Age of household is concave in post-OCP poor group families ($0.853age - 0.015age2$) implying the probability of being poor peaks at the age 28 and declines thereafter. For the pre-OCP families ($0.094Age - 0.00132Age2$) the profile is convex with the probability of being poor increasing with the age of the household head (with a minimum at 53).

In the middle and rich classes, household age profiles are convex with similar minimum which means middle and rich classes group membership probabilities increase with age after the age of household heads of approximately 28. The age profile relative to the lower-middle group is much flatter than the others, while the magnitude of the age effects is more powerful for the upper-middle income families (Figure 4).

Poor group membership probabilities increases with household size, middle and rich group membership probabilities diminish. The risk of belonging to the poor group increases by 10% for each additional member of the household. These effects are highly

Figure 4: *Age profiles - measured by the CODA coefficients - in the income groups*



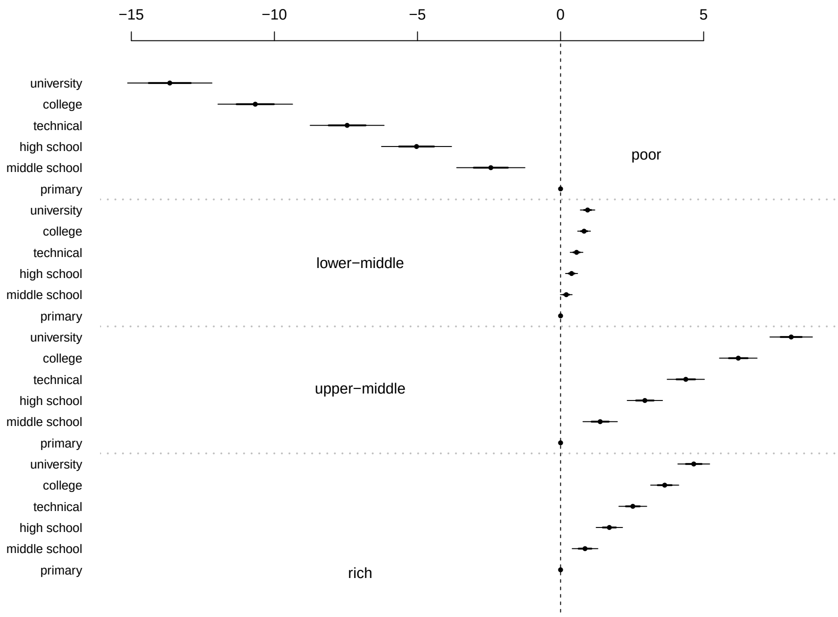
significant and are magnified for pre-OCP households.

Single parent family status results in a significantly increasing risk of poverty group membership and diminishing risk of middle and rich group memberships for the post-OCP status families. If we estimate the percentage impacts implied by this dummy variable in the regression equation of the poor group according to the Kennedy (1981) almost unbiased estimator, we find that being single parent increases the (relative) probability of being poor by 162%, while it reduces the probability of belonging to the upper-middle class by almost 100%. These effects are essentially undone for pre-OCP status families wherein single parent status appears to have no significant effects. This may be a consequence of the fact that single parent status was a relatively rare occurrence in the pre-OCP world.

Poor group membership probabilities strongly diminish with high levels of education (see Figure 5) with the effects diluted for Pre as opposed to Post OCP families. Instead, high levels of education increase the probability of being in the middle or rich groups. These effects are more pronounced for the upper-middle group membership.

For post-OCP families, the occupation of the head in collective enterprises signifi-

Figure 5: *CODA coefficients of the categories of education*

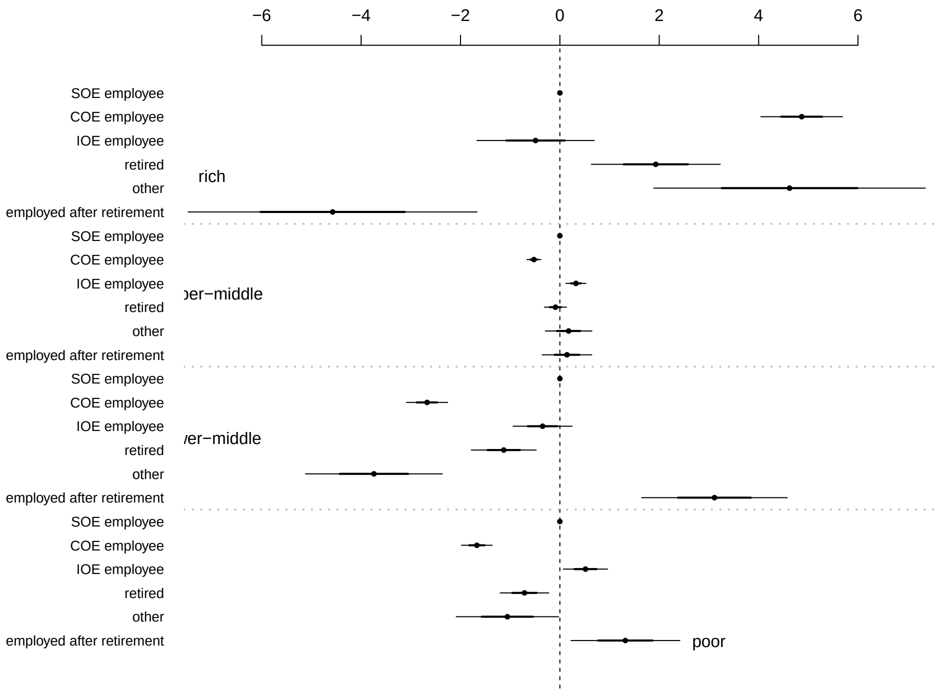


cantly increases the probability of belonging to the poor group (relative to household head employed in state owned enterprises), as well as being retired or unemployed, while being employed after retirement reduces it. These effects are not altered substantively for pre-OCP families except that the status of self-employed or employed by small scale private enterprises increases the risk of poverty group membership significantly. Largely speaking the opposite effects are observed for upper-middle and rich class membership probabilities. A notable exception is that the status of self-employed increases the probability of belonging to the rich group, while it does not make significant difference (relative to the SOE employed) in the upper-middle group (see Figure 6).

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 estab-

Figure 6: *CODA coefficients of the categories of occupation*



lish class membership. Here techniques 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 an economic variable. In the present context classes are defined by commonality in size distribution of household income which is modeled as a finite mixture of normal[9] sub distributions where each sub distribution corresponds to the size distribution for a particular sub class. A procedure for determining the number of classes was developed and parameters of the sub distributions and the class weights was estimated. 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 (means and variances of incomes) of the respective classes together with the factors that influence the probability of class status and hence class membership of individual households. These techniques have been applied in a study 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. With regard to general trends, all classes experienced real economic growth but, with the growth rates in the same rank order as the mean incomes of the respective classes, polarization between the classes is in evidence. However the population shares of the classes were changing over the sample period with the poor population growing slightly, the lower middle growing rapidly, the upper middle shrinking rapidly and the rich shrinking substantially. Note however that this is in the context of a rapidly growing overall urban population and undoubtedly the rapid growth of the lower middle class is a result of the flight from the land. Within class inequalities are growing for all groups which is consistent with Gibrat's law for describing the household income process. With regard to the factors that influence the probability of class membership, all had predictable and highly significant influences on class status probabilities. With regard to location, the provinces had very different experiences in terms of their base level and trends in class membership probabilities.

[9]Gibrat's Law provides a statistical rationale for normality of the sub distributions.

As the household matures (in terms of the age of its head) it experiences a lower chance of being in the poor group and greater chances of being in the non-poor groupings. Similarly larger household sizes are associated with higher probabilities of poor class membership as is single parenthood. Higher levels of urbanization were associated with lower probabilities of poor class membership and higher levels of education of the head of household were associated with higher probabilities of non-poor status. Employment outside of State Owned Enterprises was generally associated with lower probabilities of non-poor class status. As for the impact of the One Child Policy, households whose family size choices were completed prior to the One Child Policy had a lower chance of being in the non-poor classes.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*, Chapman and Hall, London.

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 T.W. Leo (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., I. Crawford and A. Leicester (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 George A. Slotsve (1997). "Inequality and Polarization of Male Earnings in the United States, 1968-1990", *North American Journal of Economics and Finance*, 8(2), 135–151.

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.

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.

Fry, J.M., Fry T.R.L. and Mclaren K.R. (1996). "The stochastic specification of demand share equations: restricting budget shares to the unit simplex", *Journal of Econometrics*, 73, 377–385.

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.

Kaufman, L. and Rousseeuw P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.

Kennedy, P.E. (2001). "Estimation with correctly interpreted dummy variables in semi-logarithmic equations", *American Economic Review*, 71, 4, 801.

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

Lo, Y. (2005). "Likelihood ratio tests of the number of components in a normal mixture with unequal variances", *Statistics and Probability Letters*, 71, 225–235.

Manski, C.F. (2003). *Partial Identification of Probability Distributions*, New York, Springer-Verlag.

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.

McLaren, K.R , Fry, J.M. and Fry T.R.L. (1995). "A simple nested test of the almost ideal demand system", *Empirical Economics*, 20, 149–61.

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

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 (2010). 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.

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.

A 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