Journal of Econometrics 191 (2016) 348–359

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A new approach to measuring and studying the characteristics of class membership: Examining poverty, inequality and polarization in urban China

Gordon Anderson[a,*], Alessio Farcomeni[b], Maria Grazia Pittau[c], Roberto Zelli[c]

[a] Department of Economics, University of Toronto, Canada
[b] Department of Public Health and Infectious Diseases, Sapienza University of Rome, Italy
[c] Department of Statistical Sciences, Sapienza University of Rome, Italy

ARTICLE INFO

Article history:
Available online 7 December 2015

JEL classification:
C14
I32
O1

Keywords:
Poverty frontiers
Mixture models
Class membership
Urban China

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 correlates 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.

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.[1]

Examples of disputed boundaries are not hard to find, determining the poor has probably been the most contentious (e.g. Sen, 1983; Foster, 1998). Things are not different when the focus of the analysis is on the middle or rich class (Atkinson and Brandolini, 2013; Banerjee and Duflo, 2008; Easterly, 2001; Saez and Veall, 2005). 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 reflection of societal wellness (Fitoussi et al., 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

* Corresponding author.
E-mail addresses: anderson@chass.utoronto.ca (G. Anderson), alessio.farcomeni@uniroma1.it (A. Farcomeni), grazia.pittau@uniroma1.it (M.G. Pittau), roberto.zelli@uniroma1.it (R. Zelli).

[1] More formally, randomness of class membership indicators is a direct consequence of unobserved heterogeneity. If all variables could be observed, then we would estimate class indicators, rather than probabilities. This is completely in parallel with classical mixed models in which random effect have continuous distributions (e.g., Gaussians) and instead of estimating the subject-specific values we estimate their density (e.g., through estimation of their first and second moments).

http://dx.doi.org/10.1016/j.jeconom.2015.12.006

partially reflect her/his poverty status. 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., 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 actions which differ 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 of a vector $\mathbf{x}$ of agent 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 (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 depends on the assumed nature of the process of $\mathbf{x}$. 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, $\mathbf{x}$ 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 $j = 1, \ldots, K$.

The following presents a model that simultaneously estimates class membership probabilities together with the influence of possible correlates of class membership. 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. Mixture parameters were estimated stratified by year along with the effects of individual factors governing class membership. In addition to these individual factors, role of geography, the One Child Policy and the effects of the substantial urbanization that took place over the period will be examined. 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 that 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 ($j = 1, 2, \ldots, K$), $f_j(\mathbf{x})$ 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):

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

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. Note that the covariates $\mathbf{z}_i$ are not determinants of the outcome directly but only indirectly through class membership. Obviously, each class membership refers to a latent outcome class which is conditional on the covariates. 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:

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

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 cannot be fitted directly.

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

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

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

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

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 as belonging 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 $\boldsymbol{\Psi} = \left(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_K, \boldsymbol{\xi}'\right)'$; in this case $\boldsymbol{\xi}$ consists of the elements of the component means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$ and the distinct elements of the component-covariance matrices $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K$. After estimation of the β 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.

Assume for instance that $\mathbf{x}$ is p-dimensional and $\mathbf{z}$ is q-dimensional. In the general formulation, the number of parameters is $K(p + p(p + 1)/2) + (K - 1)q$, as there will be p means and $p(p+1)/2$ parameters for the covariance matrices. Dimensionality of the parameter space generally grows quadratically with the dimensionality of $\mathbf{x}$, but only linearly with K. It should be also noted that this feature of the model can be controlled quite precisely, achieving a good balance between bias and variance, through appropriate assumptions. The dimension of the parameter space can grow linearly or even be constant as a function of p after specifying appropriate assumptions or constraints. We focus here mostly on the covariance matrices. The simple assumption of conditional independence leads to zero off-diagonal elements, so that the number of parameters is reduced to $2Kp + (K - 1)q$. The additional assumption of variance homogeneity leads to $Kp + 1 + (K - 1)q$. Other possibilities include a homogeneous correlation structure where all off-diagonal elements coincide (leading to $K(2p + 1) + (K - 1)q$), banded covariance matrices, autoregressive and similar structures. A general approach is obtained through a parameterization of the kind (Celeux and Govaert, 1995; Fraley and Raftery, 2002):

$$\Sigma_j = |\Sigma_j|^{1/p} V_j \frac{\Lambda_j}{|\Sigma_j|^{1/p}} V_j^T. \tag{5}$$

Expression (5) is based on three elements: $|\Sigma_j|^{1/p}$, V_j and $\Lambda_j/|\Sigma_j|^{1/p}$, where V_j denotes the matrix of eigenvectors and Λ_j the diagonal matrix of eigenvalues of Σ_j. These three elements control the volume, orientation and shape of the jth cluster, respectively. Assumptions can be formulated so that certain, or even all, characteristics are equal across components. For instance, if $|\Sigma_j|^{1/p}$ is constant over j, groups have the same volume. Through assumptions on V_j one can obtain components with the same orientation, or that Σ_j is diagonal. Finally, it can be assumed that all elements of Λ_j are equal, therefore having *spherical* models. Fourteen models arising from (5) are discussed in Celeux and Govaert (1995). Finally note that functional relationships between the means can also be formulated in many situations. Estimation of the model under these assumptions is straightforward and only involves minor changes (e.g., pooling of the estimates) to the M step described below.

2.2. Estimation of mixture distributions via the EM algorithm

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 $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ independently generated from (4), the log-likelihood of the parameters given the data is:

$$\ell(\boldsymbol{\Psi}|\mathbf{X}) = \sum_{i=1}^{n} \log \left(\sum_{j=1}^{K} f_j\left(\mathbf{x_i}|\mu_j, \Sigma_j\right) \cdot \frac{e^{\mathbf{z}_i \boldsymbol{\beta}_j}}{\sum_h e^{\mathbf{z}_i \boldsymbol{\beta}_h}} \right). \tag{6}$$

The maximum likelihood principle ensures that the best model of the data is the one with parameters that maximize (6). Maximizing (6) 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 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:

$$\ell_c(\boldsymbol{\Psi}|\mathbf{X}, C) = \sum_{i=1}^{n}\sum_{j=1}^{K} \mathbb{I}(C_i = j) \log\left(f_j\left(\mathbf{x_i}|\mu_j, \Sigma_j\right)\right) + \sum_{i=1}^{n}\sum_{j=1}^{K} \mathbb{I}(C_i = j) \log(w_{ij}), \tag{7}$$

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 (7) no longer involves the logarithm of sum but since C is unknown, ℓ_c cannot 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. The E-step, at the generic iteration $\nu + 1$, requires the computation of the conditional expectation of ℓ_c given the data: $\mathbb{E}_{\boldsymbol{\Psi}^{(\nu)}}\{\ell_c(\boldsymbol{\Psi}|\mathbf{X}, C)\}$. 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}:

$$\mathbb{E}_{\boldsymbol{\Psi}^{(\nu)}}\{\mathbb{I}(C_i = j)|\mathbf{X}, \mathbf{Z}_i\} = \text{prob}\{\mathbb{I}(C_i = j) = 1|\mathbf{X}, \mathbf{Z}_i\} \tag{8}$$

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

$$\tau_{ij} = \frac{w_{ij} f_j(\mathbf{x_i})}{\sum_{h=1}^{K} w_{ih} f_h(\mathbf{x_i})}, \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, K. \tag{9}$$

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 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. When $\mathbf{z}$ 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.

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),

[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.

which requires little 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. For more details on the approach outline for a slightly more general class of models, refer to Bartolucci and Farcomeni (2015). This component is directly obtained at the M step of the EM algorithm, while the second can also be obtained in closed form similarly.[3] The standard errors are then estimated as the square root of the diagonal of the expected information matrix.

2.3. Interclass comparison

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 Yitzhaki, 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 membership 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 et al., 2012b). 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 jth group and $f(\mathbf{x}_{m,j})$ is the corresponding value of the pdf at the modal point (Anderson, 2010; Anderson et al., 2012a).

What of 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 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 (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

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

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:

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

or by considering the dominance relationship between the estimated function $f_j(\mathbf{x}_i;\mu_j;\Sigma_j)$ and $f_h(\mathbf{x}_i;\mu_h;\Sigma_h)$ directly.

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 Economic Reforms, instigated in the late 1970s, appeared to promote unprecedented growth in urban incomes (as reported in Ravallion 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. The One Child Policy intervention also introduced in the late 1970s changed fundamentally the nature of the family in many respects in subsequent years. All of which affected not only the level and growth of urban disposable income but also substantially changed the way households relate to one another, one aspect of which is the extent to which households grouped and evolved into classes. Most work dealing with the Chinese income distribution is essentially based on summary statistics, which do not fully convey all the information on distributional shapes. Starting from the analysis of the whole income distribution, the approach here facilitates the identification of urban household income classes and their evolution over a decade of intense transformation in China's society.

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.[4] Unfortunately UHS is not in the public domain, so for this reason this study employed a sub-sample of ten cross-sectional annual surveys of urban households from six provinces from 1992 to 2001.[5] Among the selected provinces, three coastal and three interior provinces were selected some e.g. Guangdong, are characterized by rapid economic growth others, like Sichuan

[3] Computational details are reported in Appendix A.1.

[4] 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 8715 in 1981 to around 33,000 in 1987 and has remained about the same until the 2000s. On the quality of household income surveys in China see Bramall (2001).

[5] Many thanks are due to Benjamin et al. (2008) for providing the data which originally emanated from the National Bureau of Statistics as part of the project on Income Inequality during China's Transition which they organized.

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

k	log-lik	# parameters	BIC
2	−350049	77	700917
3	−340477	134	682380
4	−339497	191	681026
5	−339215	248	681069

are characterized by relatively slow economic growth. The selection was thought to be sufficiently representative of the regional differences. This survey has comparably defined income for all years even though the implicit value of subsidies associated with food coupons, is missing (Benjamin et al., 2008). Whilst this problem was particularly serious in the 1980s it was more or less eliminated by the 1990s, the period of study here.

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. One shortcoming of the UHS is that there is no adjustment for spatial difference in the cost of living. Since regional price differences are expected to be wide between the selected provinces, comparison across provinces cannot be implemented without employing spatial price indices (SPI) for adjusting spatial price differentials, thus Gong and Meng (2008) SPI's were employed as regional deflators.[6]

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 rule (Brady and Barber, 1948). Household incomes are reported in 1994 prices using the corresponding national deflator.

A set of household characteristics that can help in understanding class membership of each household was also extracted from the UHS dataset. The selected explanatory variables include: demographics (age and household size); employment status of household head, recoded as: State-owned enterprises employee (SOE), employee of collective enterprises (COE), SPE employee that includes small scale private enterprises employee, self-employed and employee of other types, retired, other status (unemployed, house worker, disable, student, etc.), employed after retirement (EaR); education of the household head, coded 1–7, from 1 (no schooling) to 7 (university graduates)[7]; family status (recoded as single parent or not). Province of residence (Shaanxi as reference), a time trend, and the urbanization index of the province of each year were also added.[8] A dummy variable (OCP dummy) indicating families that most likely were not involved in the one-child policy (OCP), that is households whose head was aged more than 39 year old in 1978 when the policy was introduced, was also included.

3.2. The model at work

3.2.1. Number of components and sensitivity analysis

Model (3) was estimated using size adjusted household incomes as the observable outcome assuming Gaussian component densities[9] and all the explanatory variables introduced in Section 3.1. The variables were interacted with time since they could have different impacts on the class membership in different time periods. Estimates were carried out pooling observations from all years but allowing for temporal dynamics since mixture parameters $\theta_{jt} = \left(\mu_{jt}, \sigma_{jt}\right)$ and individual class memberships w_{ijt} were stratified by year. The model was repeatedly fit from differing initial solutions. A deterministic starting solution based on separate models for the outcome based on k means, and for outcome clustering (estimated directly via Newton–Raphson) was employed. The model was then fitted 15 additional times based on random jittering of the deterministic starting solution, and then another 15 times based on random jittering of the estimates at convergence of the previous runs. The results in the empirical application were fairly stable with respect to the starting solution in the sense that the same maximum for the likelihood or a value very close to it was invariably obtained. Moreover, on simulated data the probability of observing at least once the global optimum when 30 starting solutions are used is above 99%.[10]

Given sample size of between four and five thousand observations per year, the number of components has been assessed by using the Bayesian's Information Criteria (BIC). Since BIC adds a term in the log likelihood it penalizes the complexity of the model, and in this context is particularly helpful in finding a parsimonious parameterization of the model. 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). These components can be interpreted as "poor", "lower-middle", "upper-middle" and "rich" income groups. Adding a fifth component yields a negligible improvement in fit leaving the first three components unchanged and splitting the fourth component into two classes.

A novelty of this approach is that all the parameters of the model are estimated simultaneously. A related drawback is the dependence of the classification on the specification of the explanatory variables. It is therefore important to assess the robustness to changes in the model specification due to explanatory variables entering the model. We implemented a sensitivity analysis keeping fixed only the contextual variables (provinces and urbanization index) but allowing all the individual variables to enter into each simulation step randomly with probability equal to 0.5. More precisely, at each iteration we independently flagged each predictor, with probability 0.5, to be used for modeling purposes. We then fit the model on the entire dataset, but based only on the randomly chosen subset of predictors (plus the contextual variables), for $k = 1, \ldots, 6$. BIC

6 Gong and Meng (2008) 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. As expected, high income provinces such as Guangdong are ranked as high price provinces, while low income provinces as Shaanxi have low prices. Gong and Meng (2008) 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).

7 The categorical variable of education was treated as a continuous variable in the final model since the estimated effects of education are almost perfectly linear.

8 See Appendix A.2 for details on its construction.

9 The assumption of normality may be too restrictive, since in principle any functional form can be taken into account (Pittau et al., 2010). The choice of normality stems from a threefold motivation. Firstly, mixtures of normal distributions form a 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 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, the assumption of normality, following Gibrat's law, is a result of additive shocks to the expected income of each strata.

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

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

Year	Poor			Lower-middle			Upper-middle			Rich		
	μ_t	σ_t	$\overline{w}_t\%$	μ_t	σ_t	$\overline{w}_t\%$	μ_t	σ_t	$\overline{w}_t\%$	μ_t	σ_t	$\overline{w}_t\%$
1992	839	216	9.5	1314	283	32.6	1920	394	41.8	3112	606	16.1
1993	842	223	10.0	1424	324	34.1	2288	507	41.2	3765	775	14.7
1994	866	243	10.5	1523	375	34.3	2498	589	41.4	4144	823	13.8
1995	947	236	10.9	1625	370	33.6	2565	588	41.7	4073	784	13.9
1996	1069	279	10.9	1776	435	33.5	2598	659	41.7	4229	876	13.8
1997	1027	282	11.1	1732	408	33.3	2653	628	42.1	4240	844	13.5
1998	1043	371	10.4	1777	454	31.6	2785	705	42.7	4635	1037	15.2
1999	1142	355	9.7	1918	475	30.3	3037	760	43.1	5111	1108	16.9
2000	1102	366	10.0	1971	519	30.4	3151	843	41.6	5493	1351	18.0
2001	1148	365	9.6	2106	581	31.9	3347	935	40.4	5962	1632	18.1

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

was finally used to choose the optimal number of latent classes for that iteration, that is, for the specific set of predictors selected. The "optimal" number of latent classes was selected as $k = 4$ in 822 out of 1000 replicates, while $k = 5$ was selected 178 times. The other possible values of k were never selected. The estimated mean parameters for the income groups were very stable with respect to the different specifications of the models. As expected, what varied were the individual-specific weights depending on the predictors included in the model.

3.2.2. Parameters of the components and polarization cohesion

Table 2 provides a summary of the model fits stratified by years for each class j: the estimated mean (μ_{jt}) and standard deviation (σ_{jt}) of each normal component along with its corresponding mixing proportion ($\overline{w}_{jt}$). Fig. 1 visually compares the fitted four component mixtures for some years of the analysis with the corresponding estimated kernel density.[11]

Table 2 reveals real annual income growth rates of 3.2% for the poor, 4.8% and 5.7% for the lower and upper middle class respectively, and 6.7% for the rich over the period. The marginal probability of being in the poor group (an estimate of poverty rate in this model) slightly increases at the beginning of the period (moving from 9.5% in 1992 to 11.2% in 1996) and decreases since 1996 reaching a level of 9.6% in 2001. The lower-middle income group, stable in size after the first two years, decreases moderately at the end of the period, and similarly the upper-middle group. After 1997 the rich group mixing proportions tend to increase especially in the last two years reaching the level of 18.1% in 2001. The within group inequalities grow for all groups (which is consistent with Gibrat's law).

The size of the 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 Ravallion 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 fixed a relative poverty line at 50% of the median income for each year.[12] The corresponding head count ratios are reported in Table 3.

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 50% median	Size poor group
1992	9.3	6.0	9.5
1993	9.2	9.3	10.0
1994	8.5	10.2	10.5
1995	6.2	9.5	10.9
1996	4.6	7.7	11.2
1997	5.0	8.6	11.1
1998	5.0	9.9	10.4
1999	3.4	9.9	9.7
2000	4.2	11.4	10.0
2001	3.5	11.5	9.6

The absolute poverty drastically decreases over time, while the percentage of households below the relative poverty line, which is tied to the progress of the whole distribution, exhibits a fluctuating pattern until 1997 and it increases afterwards. The size of the poor group shows values closer to the relative measure of poverty but it is more stable over time.

This mixture model has the advantage of allowing for further investigation on the progress of the poor, as well as tracking the "distance" of the poor with respect to the other classes. As previously described, memberships of each class is not determined with certainty, but each household has attached an estimated probability of belonging to each component of the mixture (see Eq. (9)). 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 Fig. 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 et al., 2012a) 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. The distance between the contiguous poor and lower-middle groups enlarges over time, and the poor–rich gap increases by more than 50%.

3.2.3. Correlate effects

The aforementioned growth in China was unevenly spread with evidence on per capita urban incomes suggesting greater advances for seaboard provinces than for inland provinces. Partly the result of regional comparative advantage, it also reflected weak government regional equalization policy, imperfect capital markets, and initial preferential policies on FDI and exports and from the growth of tax revenues as their development proceeded

[11] 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.

[12] The choice of the thresholds for measuring poverty is largely arbitrary and determines the proportion of population calculated to be at risk of poverty. Thresholds set at 50% and 60% of the national median income are those most commonly used. The 50% threshold is mostly used by the OECD and in the Luxembourg Income Study, while the 60% threshold is used as the main EU indicator of risk of poverty. Traditionally poverty in China has been measured with respect to an absolute poverty line; however relative thresholds are becoming familiar for measuring urban poverty in China. At the poverty summit on September 2013 Hong Kong announced its first official poverty line set at 50% of the median household income by household size.

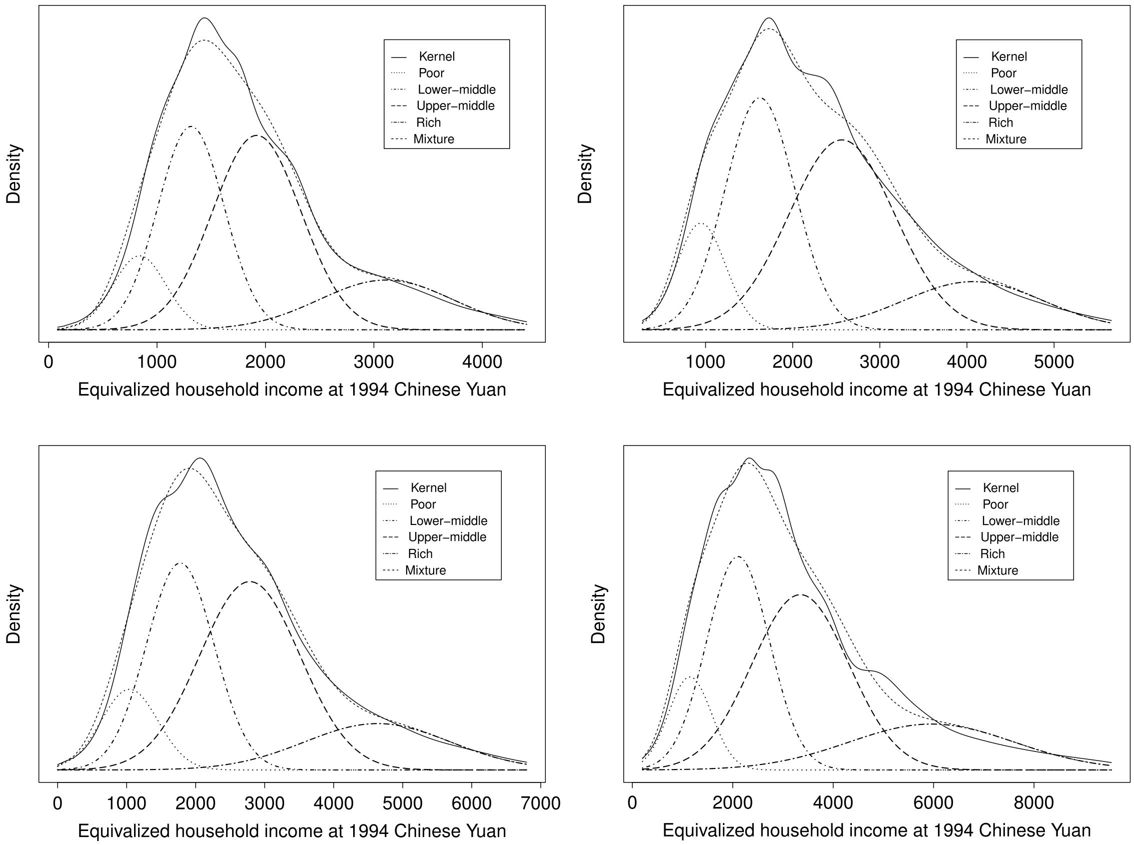

Fig. 1. Kernel density estimation and the (inflated) four-components mixture model fit, 1992, 1995, 1998 and 2001. The four components are labeled "Poor", "Lower-middle", "Upper-middle" and "Rich" according to the values of their estimated means.

Table 4
Mixture trapezoid measures. Standard errors in parenthesis.

Year	R-UM	R-LM	R-P	UM-LM	UM-P	LM-P
1992	21.637 (0.046)	35.889 (0.053)	49.268 (0.078)	13.275 (0.036)	11.220 (0.059)	12.079 (0.055)
1993	23.668 (0.047)	42.716 (0.056)	59.988 (0.085)	17.228 (0.036)	12.930 (0.063)	14.224 (0.055)
1994	24.973 (0.047)	44.222 (0.056)	64.738 (0.084)	18.057 (0.036)	13.807 (0.063)	15.175 (0.055)
1995	23.148 (0.046)	42.825 (0.055)	62.859 (0.082)	17.480 (0.037)	14.381 (0.062)	15.834 (0.055)
1996	23.665 (0.046)	39.992 (0.054)	59.034 (0.080)	14.249 (0.036)	13.937 (0.060)	15.205 (0.054)
1997	23.529 (0.047)	41.987 (0.055)	60.226 (0.079)	16.426 (0.036)	13.986 (0.059)	15.336 (0.053)
1998	25.358 (0.046)	44.459 (0.055)	59.449 (0.077)	17.009 (0.037)	13.116 (0.058)	14.473 (0.052)
1999	27.435 (0.044)	48.358 (0.054)	65.803 (0.080)	18.338 (0.037)	13.830 (0.062)	15.318 (0.055)
2000	28.800 (0.045)	49.952 (0.055)	69.612 (0.084)	18.439 (0.038)	15.031 (0.063)	16.669 (0.056)
2001	29.971 (0.046)	50.950 (0.056)	74.032 (0.091)	18.365 (0.037)	16.252 (0.066)	17.930 (0.058)

for the seaboard provinces (Anderson and Ge, 2004; Gustafsson et al., 2007). Consequently urbanization rates differed in seaboard and inland provinces with cities growing more rapidly both in size and number in the seaboard provinces (Anderson and Ge, 2005). A concomitant decline in the share in output of state owned and collective enterprise resulted in fundamental changes in the structure of employment and educational opportunities, again unevenly spread across the provinces. All of these, as factors, which are considered to influence class (Goldthorpe, 2010), are employed as correlates of class structure in urban China. At the same time the effects of the One Child Policy initiated in 1978 in changing the nature of the family, not only in its size but also in its makeup. Anderson and Leo (2009, 2013) document the increased educational investment in children and increased positive educational sorting of marriage partners in Post One Child Policy families which resulted in substantial changes in the nature of the family and increased household income dispersion[13] in urban China all of which could also be seen as influencing the circumstances of the household and thus correlate with the makeup of class. With poor as the baseline category, Table 5 contains EM estimates of the correlate effect parameters $\boldsymbol{\beta}_j$ from model (3), along with their standard errors.

Disparities between provinces (inland *versus* coastal, for example) are often regarded as the most important source of income differences, as some provinces took greater advantages than others of progressive market liberalizations and reforms during China's economic transition (Benjamin et al., 2008). Based on the estimated parameters reported in Table 5, we first focus on the potential role of locational circumstances in shaping disproportionate growth between provinces, that is unbalanced differences between class memberships. Holding all other effects constant, all provinces except Jilin present households with increasingly better odds of being in lower middle, upper middle and rich classes than in the poor class than Shaanxi, with Hubei,

13 A result of the intensified pairing of high income males and females and similarly the intensified pairing of low income males and females it increased the probability of observing poor young families or rich young families and reduced the probability of observing young middle income families.

Table 5
Estimated parameters of effects on the odds that Chinese households belong to the lower middle, upper middle and rich group relative to the poor group.

	Lower-middle		Upper-middle		Rich	
	Estimates	s.e.	Estimates	s.e.	Estimates	s.e.
Shaanxi	–		–		–	
Jilin	−0.592	0.073***	−2.740	0.078***	−3.920	0.140***
Shandong	10.712	0.206***	13.470	0.184***	16.020	0.264***
Hubei	5.162	0.135***	7.033	0.120***	8.419	0.173***
Guangdong	5.191	0.167***	7.176	0.150***	9.451	0.215***
Sichuan	5.601	0.144***	6.845	0.129***	8.189	0.185***
Time-trend	−0.268	0.025***	−0.596	0.023***	−0.873	0.036***
Jilin:trend	0.015	0.013	−0.002	0.014	−0.093	0.023***
Shandong:trend	−0.201	0.054***	−0.791	0.047***	−1.307	0.065***
Hubei:trend	0.112	0.038***	−0.367	0.033***	−0.671	0.045***
Guangdong:trend	0.740	0.049***	0.291	0.042***	−0.070	0.057
Sichuan:trend	0.069	0.042*	−0.376	0.036***	−0.721	0.050***
Age	0.081	0.007***	0.226	0.007***	0.353	0.009***
Age-squared	0.000	0.000***	−0.002	0.000***	−0.002	0.000***
Education	0.407	0.014***	0.760	0.013***	0.964	0.019***
Household size	−1.785	0.027***	−3.467	0.031***	−5.442	0.065***
Urbanization index	−4.281	0.109***	−5.415	0.098***	−7.007	0.141***
Single parent dummy	−0.831	0.054***	−1.860	0.051***	−2.776	0.072***
SOE employee	–		–		–	
COE employee	−0.850	0.027***	−1.557	0.029***	−2.420	0.060***
SPE employee	−0.531	0.040***	−0.741	0.042***	−0.720	0.065***
Retired	0.072	0.045	0.041	0.042	−0.118	0.060**
Other Status	−2.065	0.096***	−3.696	0.135***	−4.645	0.247***
EaR	2.873	0.136***	4.468	0.094***	5.018	0.116***
OCP dummy	−1.092	0.143***	−0.686	0.134***	−0.795	0.209***
Education:OCP	−0.062	0.018***	−0.173	0.015***	−0.040	0.020**
Household size:OCP	0.310	0.027***	0.414	0.030***	0.442	0.061***
Urb.index:OCP	−0.590	0.040***	−0.844	0.035***	−0.855	0.047***
Single parent:OCP	1.902	0.097***	2.403	0.098***	3.108	0.144***
COE employee:OCP	1.664	0.181***	1.129	0.213***	1.558	0.323***
SPE employee:OCP	0.449	0.272*	−0.494	0.286*	−3.139	0.562***
Retired:OCP	0.034	0.081	−0.468	0.072***	−0.738	0.097***
Other status:OCP	−0.278	0.192	−1.667	0.244***	−3.374	0.419***
EaR:OCP	−1.192	0.175***	−2.010	0.124***	−2.450	0.153***
Education:trend	0.025	0.002***	0.049	0.002***	0.067	0.003***
Household size:trend	0.041	0.004***	0.041	0.005***	0.038	0.009***
Urb.index:trend	0.018	0.028	0.338	0.024***	0.643	0.034***
OCP:trend	0.065	0.010***	0.103	0.009***	0.037	0.012***
Intercept	6.380	0.198***	7.582	0.191***	8.797	0.301***

* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Guangdong and Sichuan holding very similar advantage and Shandong holding by far the greatest advantage. Those odds are trending downwards for the rich class and upward for the lower middle class. Jilin presents consistently lower odds than Shaanxi of being in the higher classes and those odds are not changing substantially over time. As for the One Child Policy, holding all else constant, families formed after the OCP have increased odds of belonging to higher income groups as is to be expected given the increased positive assortative matching and investment in education. To lend support in interpreting the correlate effects we estimate the *average expected probability* of belonging to a certain income class associated with different values of one input variable holding other inputs fixed.[14] Particularly, letting $\mathbf{z} = (u, \mathbf{v})$ with u the input of interest and $\mathbf{v}$ the vector of the other correlates, we evaluate the predictive change in expected probability of household i ($i = 1, \ldots, n$) belonging to class j for the input of interest u evaluated at two different values, say u_{lo} and u_{hi} defined as follows:

$$\delta_i = \mathrm{E}(\pi_{ij}|u^{(hi)}, \mathbf{v}_i, \hat{\theta}) - \mathrm{E}(\pi_{ij}|u^{(lo)}, \mathbf{v}_i, \hat{\theta})$$

where $\mathrm{E}(\pi_{ij}) = \widehat{\pi_{ij}}$ represents the estimated probability of household i to belong to class j with $\hat{\theta}$ the vector of parameters as estimated in model (3). The average across the n observations of these predictive comparisons δ_i as well as the average of the expected probabilities depends on the actual distribution of the other inputs and does not rely on an arbitrary choice of references.

The estimated average probabilities of belonging to each income class, assuming that households differ only by the province of residence are reported in Table 6 for each year of the analysis. Table 6 corroborates the previous general view: the estimated probabilities of being poor for households residing in Jilin are always higher than the probabilities of being poor for Shaanxi's families and this pattern tends to be more pronounced over time. On the opposite side, the estimated probabilities of being rich in 1992 are 0.036 and 0.148 for households residing in Jilin and in Shaanxi respectively, with an average estimated difference equal to 0.11. This estimated difference increases over time and in 2001 reaches the value of 0.39, indicating an increasing divergence

[14] This way of summarizing the structure of the predictive model follows the approach proposed by Gelman and Pardoe (2007). A rough procedure that is often used is to evaluate changes around a "central value", usually fixed at the mean or at the median of the data. With many predictors and interactions single central values are not necessarily representative of the entire distribution especially with categorical inputs for which the concept of "central value" becomes less meaningful. Moreover, since our model is not linear the choice of baselines for evaluating changes in probabilities is quite arbitrary. We thank an anonymous referee for pointed this out.

Table 6
Average predicted probabilities of being in the poor, lower-middle, upper-middle and rich income classes among households differing in province of residence over time.

Poor	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Jilin	0.615	0.647	0.669	0.677	0.686	0.689	0.677	0.653	0.622	0.579
Shandong	0.001	0.002	0.004	0.005	0.007	0.010	0.010	0.009	0.009	0.008
Hubei	0.055	0.072	0.088	0.099	0.106	0.110	0.097	0.078	0.063	0.047
Guangdong	0.035	0.031	0.026	0.019	0.013	0.009	0.004	0.002	0.001	0.000
Sichuan	0.045	0.062	0.078	0.091	0.101	0.107	0.098	0.081	0.066	0.051
Shaanxi	0.514	0.547	0.564	0.562	0.558	0.541	0.484	0.411	0.343	0.276
Lower-middle										
Jilin	0.333	0.302	0.278	0.263	0.249	0.241	0.232	0.226	0.218	0.219
Shandong	0.293	0.350	0.396	0.443	0.491	0.535	0.555	0.577	0.614	0.659
Hubei	0.403	0.428	0.443	0.456	0.469	0.482	0.480	0.485	0.506	0.535
Guangdong	0.346	0.385	0.411	0.435	0.457	0.474	0.467	0.461	0.479	0.502
Sichuan	0.499	0.524	0.538	0.547	0.555	0.563	0.556	0.557	0.572	0.599
Shaanxi	0.264	0.221	0.193	0.175	0.154	0.141	0.127	0.117	0.112	0.112
Upper-middle										
Jilin	0.016	0.016	0.017	0.017	0.018	0.021	0.025	0.031	0.034	0.040
Shandong	0.183	0.185	0.193	0.188	0.185	0.178	0.169	0.159	0.138	0.123
Hubei	0.192	0.175	0.171	0.158	0.151	0.143	0.135	0.126	0.109	0.097
Guangdong	0.154	0.151	0.157	0.154	0.153	0.151	0.143	0.134	0.116	0.106
Sichuan	0.152	0.140	0.138	0.130	0.127	0.124	0.122	0.120	0.107	0.099
Shaanxi	0.074	0.071	0.072	0.070	0.072	0.074	0.075	0.075	0.069	0.064
Rich										
Jilin	0.036	0.036	0.037	0.044	0.046	0.049	0.066	0.090	0.126	0.162
Shandong	0.524	0.464	0.407	0.363	0.317	0.277	0.266	0.255	0.239	0.210
Hubei	0.351	0.324	0.298	0.288	0.274	0.266	0.288	0.311	0.323	0.320
Guangdong	0.466	0.433	0.406	0.393	0.377	0.366	0.386	0.403	0.404	0.392
Sichuan	0.304	0.274	0.245	0.233	0.218	0.207	0.224	0.242	0.255	0.251
Shaanxi	0.148	0.161	0.170	0.193	0.215	0.244	0.314	0.398	0.476	0.548

between Jilin and Shaanxi. For the remaining four provinces the estimated average probabilities of being poor are substantially smaller with an almost stable pattern over the entire period except for residents in Guangdong for whom the probability of being poor at the end of the period is almost zero. Shandong presents households that on average increase constantly over the period their probabilities of belonging to the lower-middle income group. However, the probabilities of belonging to the upper-middle and to the rich groups exhibit a clear downward trend, more pronounced for the rich group. Hubei, Guangdong and Sichuan display households characterized by an increasing trend in belonging to the lower-middle class, while their average probabilities to belong to the upper-middle and especially to the rich group decrease over time. Overall, although there are only few representative provinces, the "locational premium" (Milanovic, 2015) is large but decreasing: in fact we observe with few exceptions, a steady erosion of the contribution of geography in forming all the classes, especially for the poor and the rich components.

Increases in the age of the head of the household, effectively a measure of family vintage, increases the likelihood of being in the higher classes but at a diminishing rate. Household size increases the chance of being in the poor class over the other classes, though these effects are trending downward and are uniformly smaller for pre OCP families. With respect to provincial urbanization rates, holding other effects constant, the odds of Chinese families of belonging to groups other 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 family's transit from rural to urban China was entry into the poor class initially swelling its numbers relative to the other classes. Single parenthood increases the chance of being in the poor group for post OCP families and reduces it for pre OCP families.

The results underscore the important role played by education in determining the progress of the income distribution: higher levels of education of the head of household in fact increase the chances of being in the upper classes but the effect is diminished for pre OCP families. On average, households whose head has a low level of education are more likely to belong to the poor income class, while chances of being poor are very low for those with higher education. This pattern is even more appreciable when we look at the average predictive probability plot (Fig. 2). With respect to high educated individuals (university), in 1992, being low educated (primary school) translates into more than 7 percentage points in the likelihood of being poor, and more than 13 points in 2001. More variation characterizes the patterns of households belonging to the rich group. In 1992 the estimated predictive differences between primary school and university is 0.19. Since then, the gap steadily increases, peaking at 0.36 in 2001 (Fig. 3). Results in the figure indicate that increasing returns to university-level education are the main cause in explaining this growing gap. In fact, it is also persistent the difference between holding an university degree or a high school diploma: the expected gap moves from 0.10 in 1992 to 0.20 in 2001.

Regarding employment status, we do not detect any significant change in its association with income classes over time. Over the whole period of analysis, being employed in a collective enterprise (COE) reduces the likelihood of being in the upper income classes with respect to households' head working in a State-Owned enterprise (SOE). Similarly, but with effects diluted, for self-employed or employed in private enterprises (SPE). As expected, households whose head is unemployed or house worker present higher probabilities of being poor. Retired households' heads, holding other effects constant, have a greater chance of being in the middle income classes but not the rich class. Only when the retired individuals keep on working (EaR), does the chance of being rich increase. For pre OCP families the odds of belonging to income groups other than to the poor are essentially the same for SOE and COE employees, and significantly lower for SPE employees. This evidence shows that aged families' heads suffer the most from the reconstruction and privatization of many middle and small sized State-Owned enterprises in the 1990s.

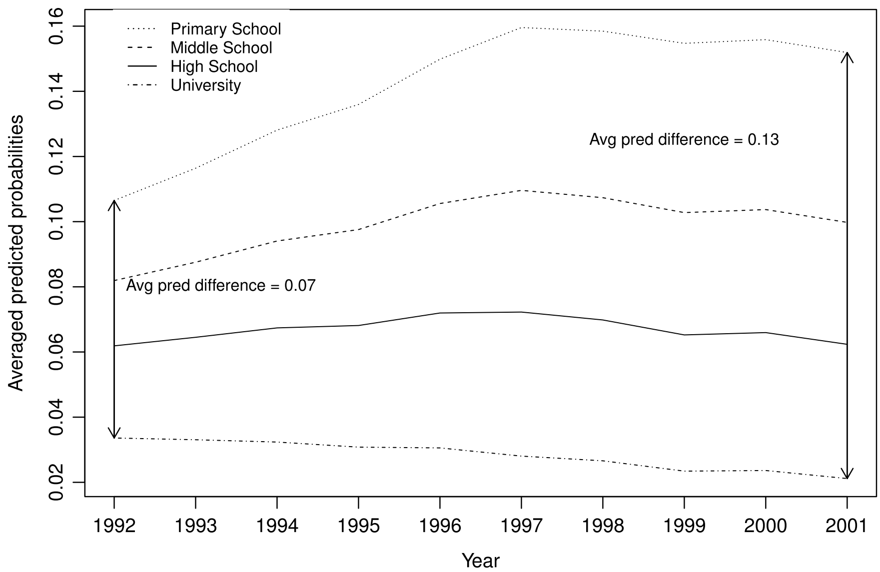

Fig. 2. Average predictive difference in probability of being poor over time, comparing individuals with different levels of education: estimated difference between primary school and university ranges from 0.07 in 1992 to 0.13 in 2001.

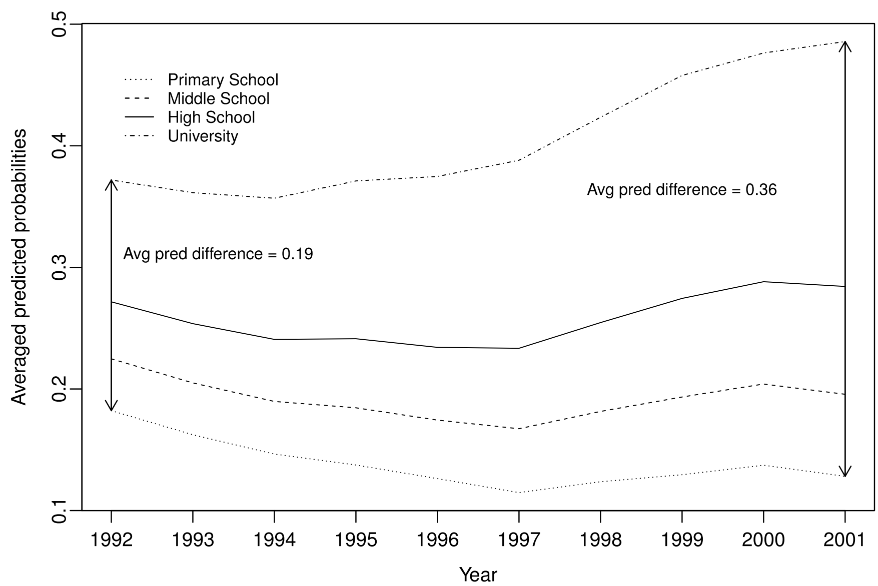

Fig. 3. Average predictive difference in probability of being rich over time, comparing individuals with different levels of education: estimated difference between university and primary school ranges from 0.19 in 1992 to 0.36 in 2001.

With the poor group as the reference category, the estimated odds (relative to SOE employees) for occupation categories are shown in Fig. 4 for pre and post OCP households.

4. Concluding remarks

There has been a long standing tradition in economics of classifying agents into groups in order to study their collective wellbeing, usually this involves the defining an artificial boundary or frontier to establish class membership. Here a new technique has 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 over the period 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.

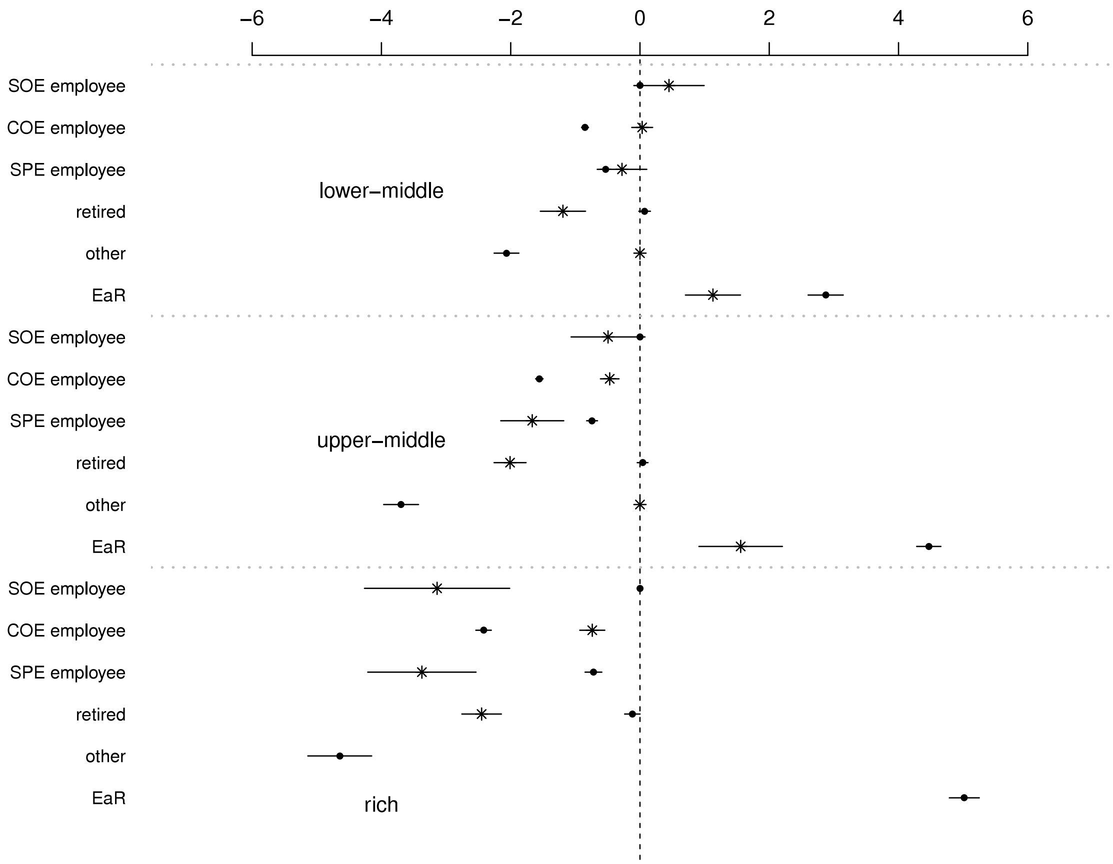

Fig. 4. Estimated log odds ratios with relative ±2 standard errors for occupation categories: post OCP families (full dots) and pre OCP families (stars). Poor group is the reference category.

Class membership was significantly influenced by all the factors considered in a predictable fashion. In particular, the rapid rate of urbanization, the locational and skill *premia* and enterprise reform, that occurred in the state sectors in the 1990s and the One Child Policy all had profound effects upon class membership probabilities and some of these effects become more important over time.

Appendix

A.1. Expected information matrix through Oakes identity

Let for ease of notation

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

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

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

evaluated at the MLE. The algebra is not simple (see also Bartolucci and Farcomeni,2015), but the rationale is easily understood: the information available can be seen as the difference between the *complete* information, and the *missing* information. The complete information is the information associated with the observed data and the latent classes, if those were observed. The missing information is the information associated with the latent classes.

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 vjth 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 β vector belonging to the same component of the mixture. This is directly obtained through first and second derivatives of the complete log-likelihood.

Let now for ease of notation

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

With similar algebra, it can be seen that the second summand is also given by a block-diagonal matrix, with

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

A.2. 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

References

Agresti, A., 2002. Categorical Data Analysis. Wiley, New York.
Alkire, S., Foster, J.E., 2011. Understandings and misunderstandings of multidimensional poverty measurement. J. Econ. Inequal. 9, 289–314.
Anderson, G.J., 2010. Polarization of the poor: Multivariate relative poverty measurement sans frontiers. Rev. Income Wealth 56 (1), 84–101.
Anderson, G.J., Crawford, I., Leicester, A., 2011. Welfare rankings for multivariate data: A non-parametric approach. J. Public Econ. 95 (3-4), 247–252.
Anderson, G.J., Ge, Y., 2004. Do economic reforms accelerate urban growth: The case of China. Urban Stud. 11, 2197–2210.
Anderson, G.J., Ge, Y., 2005. The size distribution of Chinese cities. Reg. Sci. Urban Econ. 35, 756–776.
Anderson, G.J., 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. Rev. Income Wealth 55, 607–629.
Anderson, G.J., Leo, T.W., 2013. An empirical examination of matching theories: The one child policy, partner choice and matching intensity in urban China. J. Comp. Econ. 41 (2), 468–489.
Anderson, G.J., Linton, O., Leo, T.W., 2012a. A polarization-cohesion perspective on cross country convergence. J. Econ. 17, 49–69.
Anderson, G., Linton, O., Whang, Y.J., 2012b. Nonparametric estimation and inference about the overlap of two distributions. J. Econometrics 171 (1), 1–23.
Atkinson, A.B., Brandolini, A., 2013. On the identification of the middle class. In: Gornick, J.C., Jäntti, M. (Eds.), Income Inequality. Stanford University Press, pp. 77–100.
Banerjee, A.V., Duflo, E., 2008. What is middle class about the middle classes around the world? J. Econ. Perspect. 22 (2), 3–28.
Bartolucci, F., Farcomeni, A., 2015. Information matrix for hidden Markov models with covariates. Stat. Comput. 25, 515–526.
Benjamin, D., Brandt, L., Giles, G., Wang, S., 2008. Income inequality during China's economic transition. In: Brandt, L., Rawski, T. (Eds.), China's Great Economic Transformation. Cambridge University Press, New York, pp. 729–775 (Chapter 18).
Brady, D.S., Barber, H.A., 1948. The pattern of food expenditures. Rev. Econ. Stat. 30, 198–206.
Bramall, C., 2001. The quality of China's household income survey. China Q. 167, 689–705.
Brandt, L., Holz, C.A., 2006. Spatial price differences in China: Estimates and implications. Econ. Dev. Cult. Change 55 (1), 43–86.
Celeux, G., Govaert, G., 1995. Gaussian parsimonious clustering models. Pattern Recognit. 28, 781–793.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via EM algorithm. J. R. Stat. Soc. B Stat. Methodol. 69, 1–38.
Easterly, W., 2001. The middle class consensus and economic development. J. Econ. Growth 6 (4), 317–335.
Fang, C., Wailes, E., 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. Stat. Comput. 22, 141–152.
Fitoussi, J.-P., Sen, A., Stiglitz, J.E., 2011. Mismeasuring Our Lives: Why GDP Doesn't Add Up. The New Press.
Foster, J.E., 1998. Absolute versus relative poverty. Amer. Econ. Rev. 88 (2), 335–341. Papers and Proceedings.
Foster, J.E., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 42, 761–766.
Fraley, C., Raftery, A.E., 2002. Model based clustering, discriminant analysis, and density estimation. J. Amer. Statist. Assoc. 97, 611–631.
Gelman, A., Pardoe, I., 2007. Average predictive comparisons for models with nonlinearity, interactions, and variance components. Sociol. Methodol. 37, 23–51.
Goldthorpe, J.H., 2010. Analysing social inequality: A critique of two recent contributions from economics and epidemiology. Eur. Sociol. Rev. 26 (6), 731–744.
Gong, C.H., Meng, X., 2008. Regional Price Differences in Urban China 1986–2001: Estimation and Implication. IZA Discussion Paper n. 3621.
Grusky, D., Kanbur, R. (Eds.), 2006. Poverty and Inequality. Stanford University Press, Stanford, California.
Gustafsson, B., Li, S., Sicular, T. (Eds.), 2007. Inequality and Public Policy in China. Cambridge University Press, New York and Cambridge.
Hamilton, B.W., 2001. Using Engel's law to estimate CPI bias. Amer. Econ. Rev. 91 (3), 619–630.
Kakwani, N., Silber, J., 2008. Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave Macmillan.
Keribin, C., 2000. Consistent estimation of the order of mixture model. Sankhyā 62, 49–66.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Ann. Statist. 20, 712–736.
McLachlan, G., Peel, D., 2000. Finite Mixture Models. Wiley, New York.
Melnykov, V., Maitra, R., 2010. Finite mixture models and model-based clustering. Stat. Surv. 4, 80–116.
Milanovic, B., 2015. Global inequality of opportunity: how much of our income is determined by where we live. Rev. Econ. Stat. Published on line: http://dx.doi.org/10.1162/REST_a_00432.
Nussbaum, M.C., 2011. Creating Capabilities: The Human Development Approach. Harvard University Press.
Oakes, D., 1999. Direct calculation of the information matrix via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 61, 479–482.
Pittau, M.G., Zelli, R., Johnson, P.A., 2010. Mixture models, convergence clubs and polarization. Rev. Income 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., Chen, S., 2007. China's (uneven) progress against poverty. J. Dev. Econ. 82 (1), 1–42.
Saez, E., Veall, M.R., 2005. The evolution of high incomes in Northern America: Lessons from Canadian evidence. Amer. Econ. Rev. 95 (3), 831–849.
Sen, A.K., 1983. Poor relatively speaking. Oxf. Econ. Pap. 35, 153–169.
Sutton, J., 1997. Gibrat's legacy. J. Econ. Lit. 35, 40–59.
Tao Yang, D., 1999. Urban-biased policies and rising income inequality in China. Amer. Econ. Rev. 89 (2), 306–310. Papers and Proceedings.
Wu, X., Perloff, J.M., 2005. China's income distribution, 1985–2001. Rev. Econ. Stat. 87, 763–775.
Yitzhaki, S., 1994. Economic distance and overlapping of distributions. J. Econometrics 61, 147–159.