Mixture models: Identifying consumption classes in post-liberalization India

Sudeshna Maitra, York University, Toronto*

January 6, 2022

Abstract

A mixture model is a probabilistic model that allows us to make inferences about the characteristics of sub-populations from observations on the overall population, without any information about the membership of individuals in the sub-populations or even the number of sub-populations. In this chapter, I present the theory of mixture models, and an application in which I identify consumption classes in urban India in 1999-00 (NSS). Suppose there are 3 sub-populations – a lower, a middle and an upper consumption class – determined by the total number of different durables owned by households. I construct a three-component (or three-class) mixture model of household durable ownership, which is assumed to be distributed binomially by class. I then demonstrate the use of the Expectation Maximization (EM) algorithm to estimate the size and mean durables owned by each class, as well as the probability that a household with a given number of durables belongs to a given class. Finally, I show how to assign households to classes using the mixture estimates, which allows further investigation of class-specific characteristics.

1 Introduction

A mixture model is a probabilistic model that allows us to make inferences about the characteristics of sub-populations from observations on the overall population, without any information about the membership of individuals in the sub-populations or even the number of sub-populations. In this chapter, I present an introduction to mixture models in a specific application – the identification of

*Dept. of Economics, 4700 Keele Street, Toronto, ON M3J 1P3, Canada. Email: maitra@yorku.ca

urban consumption classes (or sub-populations) in India after the widespread liberalization policies of 1991 came into effect.

Suppose that in the overall urban population, there are 3 sub-populations – a lower, a middle and an upper class – determined by the total number of different durables owned by households. Can we identify the size and the (distinct) durable ownership pattern of each class – without specifically knowing the class-membership of households – such that these, in combination, can generate the empirically observed durable ownership pattern in the entire urban poulation? Supposing that we could indeed identify classes in this way, how can we tell how many classes (or sub-populations) there are? How can we be sure that sub-populations do exist in the first place?

A finite mixture model of durable ownership provides an intuitive framework for answering the questions posed above. In the following sections, I construct a three-component (or three-class) mixture model of household durable ownership, which is assumed to be distributed binomially by class. I then demonstrate the use of the Expectation Maximization (EM) algorithm (using data from the Indian National Sample Survey, urban sub-sample, 1999-00) to estimate the size of each class and the mean number of durables owned by each class, as well as the probability that a household with a given number of durables belongs to a given class. Also, I point to challenges that may arise in the interpretation of mixture estimates in general, and how to address them in our application so as to generate informative and meaningful estimates.

2 Finite Mixture Models

Consider a population of households where characteristic Y is distributed according to the density function $f(y)$.[1] Suppose now that there are n sub-populations in a population of households (where n is finite), and the characteristic Y in households from sub-population i follow the distribution $f_i(y)$, for each $i = 1, 2, ..., n$. Then,

$$f(y) = \sum_{i=1}^{n} \theta_i f_i(y) \tag{1}$$

where θ_i is the probability that any household belongs to sub-population i. Since households must belong to one of the n sub-populations, $\sum_{i=1}^{n} \theta_i = 1$. θ_i are referred to as mixing probabilities. Equation

[1] A note on notation: throughout the paper, capital letters (e.g. Y) denote a variable that can take several values, while small letters (e.g. y) denote specific values. The only exception to this rule is the label N, which is used to denote sample size.

(1) represents an $n-$component mixture model (McLachlan and Peel (2000)).[2]

Note that the actual class membership of households is unknown. Let J denote the latent variable that represents the (unknown) class to which any household belongs, $J = 1, 2, ..., n$. Suppose that the joint distribution of characteristics Y and sub-population membership J is given by $\rho(y, j)$. Then $f(y)$ in (1) above clearly represents the marginal distribution of Y derivable from $\rho(y, j)$. Likewise, $f_i(y)$ represents the conditional distribution, or, the distribution of Y conditional on belonging to sub-population i. More explicitly, in the mixture setup:[3]

$$\rho(y, i) = \theta_i f_i(y) \tag{2}$$

for any $i = 1, 2, ..., n$, and

$$f(y) = \sum_{i=1}^{n} \rho(y, i) = \sum_{i=1}^{n} \theta_i f_i(y) \tag{3}$$

Additionally, using Bayes' Law, the conditional probability that any household with characteristic y belongs to class j is given by:

$$Pr(J = j/Y = y) = \frac{\theta_j f_j(y)}{\sum_{i=1}^{n} \theta_i f_i(y)} \tag{4}$$

for any $j = 1, 2, ..., n$.

Let us look at some numerical examples of mixture models such as in (1). For specificity, suppose $n = 2$ (viz. a two-component mixture model) and suppose that the sub-population densities $f_i(y; \mu_i, \sigma^2)$ are normal with parameters (μ_i, σ^2). What would the marginal density $f(y)$ look like for different mixing probabilities $\theta_1, 1 - \theta_1$? The graphs in Figure 1, taken from McLachlan and Peel (2000), provide a visual illustration of what the marginal density looks like when it is a mixture of two normal densities.

[INSERT FIGURE 1(a) AND FIGURE 1(b)]

[2] In (1), $f(y)$ and $f_i(y)$ refer to univariate distributions of a single characteristic Y. Mixture models may also apply to multivariate distributions $f(y_1, y_2, ..., y_k)$ and $f_i(y_1, y_2, ..., y_k)$ of k characteristics $(Y_1, Y_2, ..., Y_k)$. See McLachlan and Peel (2000).

[3] Recall the relationship between joint and marginal distributions: if $f(a, b)$ represents the joint probability of $A = a$ and $B = b$, then the marginal probability that $A = a$ is $\Pr(A = a) = \sum_x f(a, x)$ (where x represents all possible values that B can take).

Fig. 1(a): Plot of a mixture density of 2 univariate normal sub-populations (components) in equal proportions with common variance 1, and means 0 and (i) 1; (ii) 2; (iii) 3; (iv) 4.

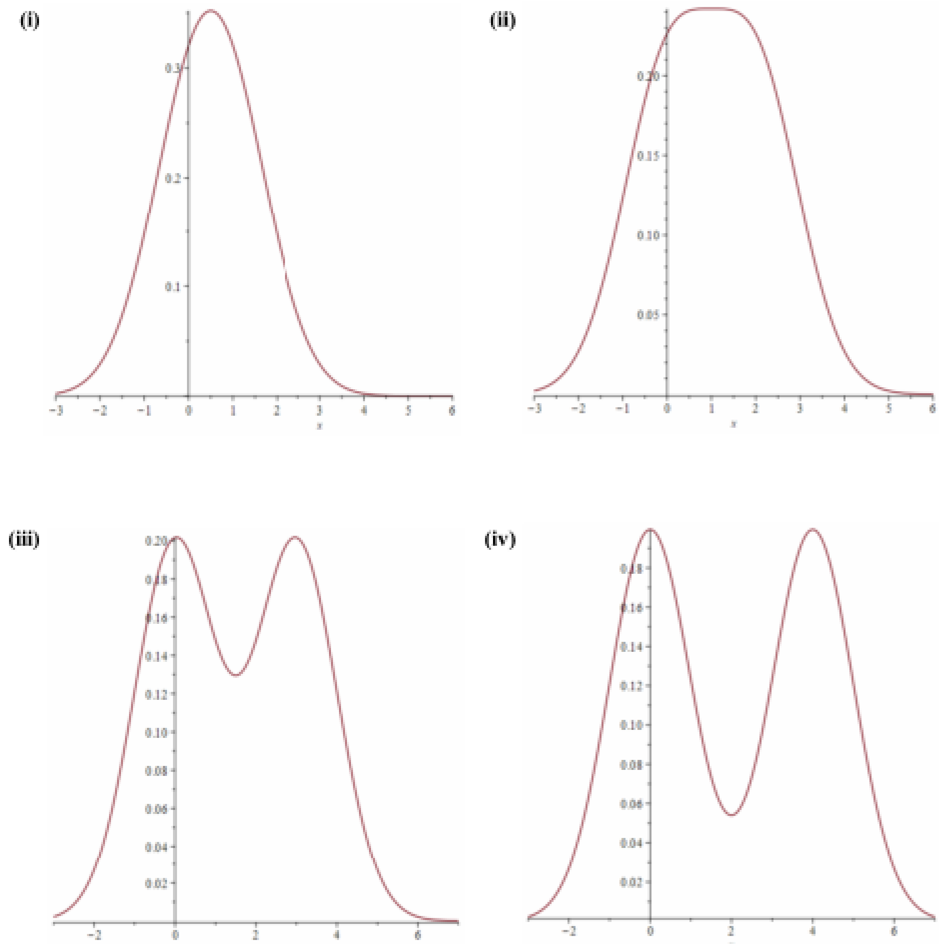


Fig. 1(b): Plot of a mixture density of 2 univariate normal sub-populations (components) in proportions 0.75 and 0.25 with common variance 1, and means 0 and (i) 1; (ii) 2; (iii); (iv) 4.

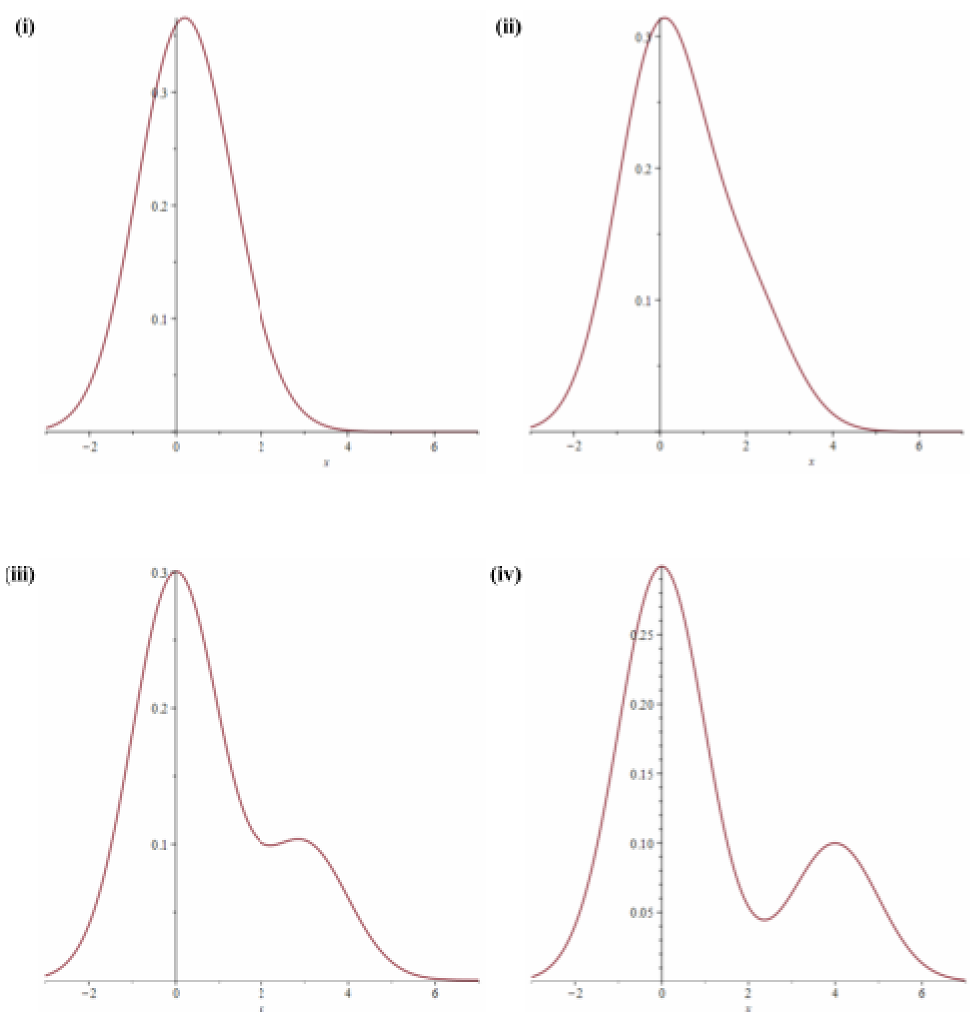


Figure 1(a) plots the (marginal) density of a mixture of 2 components mixed in equal proportion. The densities of the 2 components are normally distributed with common variance 1. The mean of the first component is 0 throughout, while the mean of the second component is 1, 2, 3 and 4, respectively, in diagrams (i)-(iv). In Figure 1(b), the components are the same as above but they are mixed in proportions 0.75 (θ_1) and 0.25 (θ_2), respectively, instead of equal shares. Note how the shapes of the marginal densities are bimodal if the 2 components have sufficient separation (in terms of their means) between them. Note also (as in Figure 1(b)(i)-(ii)), how an asymmetric density may be obtained by mixing 2 symmetric components, e.g. when the means of the component densities are close enough and the distinction between them is obscured by unequal mixing probabilities. Varying the number of components, means, variances and mixing probabilities can therefore generate a large range of shapes of the marginal density.

In Figures 1, however, we have constructed the marginal distribution $f(y)$ by mixing two known component distributions $f_i(y; \mu_i, \sigma^2)$, that are normal. In empirical applications, the situation is reversed – we have at our disposal, an empirical distribution of y sampled from the population distribution $f(y)$, and we seek to determine the size and characteristics of sub-populations (if any) that may have mixed to generate it. Given the flexibility of shapes that we have demonstrated can be obtained by mixtures, we could of course "fit" a mixture model to the data (using methods described in the sections below) assuming the existence of a certain number of sub-populations. But can we argue that sub-populations exist simply because they are able to "explain" our data? Consider, for example, the Pickering-Platt debate in the 1950s and 1960s about the patho-physiology of hypertension (McLachlan and Peel (2000)). Platt (1963) claimed that the distribution of blood pressure was a two-component mixture of "hypertensive" and"normotensive" sub-populations. But Pickering (1968) argued that "hypertensive" was merely a label assigned to those with blood pressure readings in the upper tail of the distribution for a single population, viz. that the blood pressure distribution was not a mixture at all!

How then do we determine if there do indeed exist sub-populations that have mixed to generate our data? Furthermore, how many sub-populations (if any) exist in the mixture? In the next section, we discuss these issues in the context of identifying consumption classes in post-liberalization India.

3 Consumption classes in post-liberalization India: Data and Definitions

What happened to the lower, middle and upper class in urban India after enactment of the liberalization policies of 1991? Did they become more affluent or stagnate economically as the liberalization policies came into effect? Did the relative proportions (or sizes) of the different classes in the urban population undergo a change?

To answer any of these questions, we would first have to determine *who* constitutes the lower, middle and upper classes, based on their level of affluence and income. One way to do this would be to assume cutoff levels of income or household expenditure that we believe might correspond to the different classes (Banerjee and Duflo (2008); Birdsall et al (2000); Ravallion (2010); Birdsall (2010); Ablett et al (2007); IBEF (2005)). Another approach would be to define the middle class to consist of every household that lies between, say, the 20^{th} and 80^{th} percentile of income in the economy (Easterly

(2001)). As we can see, these approaches are quite sensitive to the class-specific cutoffs being used. In addition, defining the middle class to lie between the 20^{th} and 80^{th} percentile implies that the size of the middle class is always 60% of the population, with no room to change. Furthermore, while cutoffs-based approaches implicitly assume clean, non-overlapping ranges of incomes/expenditures of the different classes, the general lack of consensus among researchers on specific cutoff levels suggests that it might be more appropriate to think of class boundaries in a probabilistic sense, i.e. by thinking of income/expenditure ranges corresponding to different classes as being overlapping instead of being mutually exclusive.[4][5]

In addition to the larger problems associated with cutoffs-based approaches, there is a specific problem surrounding the measurement of household expenditures from the Indian National Sample Survey (NSS), 1999-00 survey. The recall period used for reporting expenditures in the 1999-00 NSS questionnaire was shorter than in previous years (Deaton and Kozel (2005)). This could have resulted in higher total household expenditures reported in 1999-00 than in previous years, not due to any real increase in consumption but due to people being able to recall their more recent purchases with greater accuracy! To avoid running into this issue with measures of expenditure – especially when we compare our findings across previous rounds of NSS – we turn to a different measure of consumption to define classes, viz. durable ownership. Our measure of durable ownership is drawn from a survey question that asks respondents if a certain durable good is present in the household at the time of the survey. This measure is immune to the change in recall period, making our analysis robust to comparisons over the 1990's (Section 5.2).

Apart from the general ease (and reliability) of measurement of durable ownership, it is arguably also a good variable for measuring consumption standards of households over time. This is because durables are a store of utility that represent the stock component of household wealth rather than the flow component embodied in PCE. The ownership of durables assures the realization of a stream of consumption utility in future periods. These characteristics make durables ownership a useful measure of consumption *standards* – a permanent, sustainable aspect of consumption (Bar-Ilan and Blinder (1988)) – and, hence, an appropriate choice of variable for the identification of consumption classes.[6]

[4]See Section 6, where we discuss how a traditional clustering model such as the $K-$Means model would lead to a non-probabilistic assignments of households to classes.

[5]As Ravallion (2010) points out, some of the disagreement about cutoffs is due to the level of income of the countries concerned, viz. whether a developed or a developing economy. However, there does not appear to be a clear consensus on cutoffs even in studies that focus solely on developing nations.

[6]The approach adopted in this paper may, therefore, be considered a 'dual' approach to the expenditure-cutoffs-based approach – here, durables ownership is used to identify the classes and the expenditure-ranges of the classes thus identied are subsequently explored (instead of using expenditures to identify classes in the first place). In addition, the mixture

Our goal, therefore, is to derive a definition of classes – lower, middle and upper – in urban India from the 55th Round of the National Sample Survey data (1999-00), using data on the durable ownership of households. Our underlying premise is only that higher classes enjoy a higher standard of living (or, are more affluent) than lower classes; no further specifications are made about who or what constitutes each class. In particular, we use the ownership of different durable goods by households to be a measure of the standard of living enjoyed therein.

We consider 12 specific durable goods in our analysis, as these have often been linked to middle class status in India in existing literature (Banerjee and Duflo (2008), Senauer and Goetz (2003)) – viz. record player/gramophone, tape/CD player, radio, television, VCR/VCP, electric fan, washer, refrigerator, air conditioner, bicycle, motor cycle/scooter and motor car/jeep. We assume that owning a higher number of different durable goods indicates a higher standard of living. Hence, our variable of interest – call it Y – is the total number (out of 12) of different durable goods owned by households as reported in the urban subsample NSS, 1999-00 ($Y = 0, 1, 2, ..., 12$).

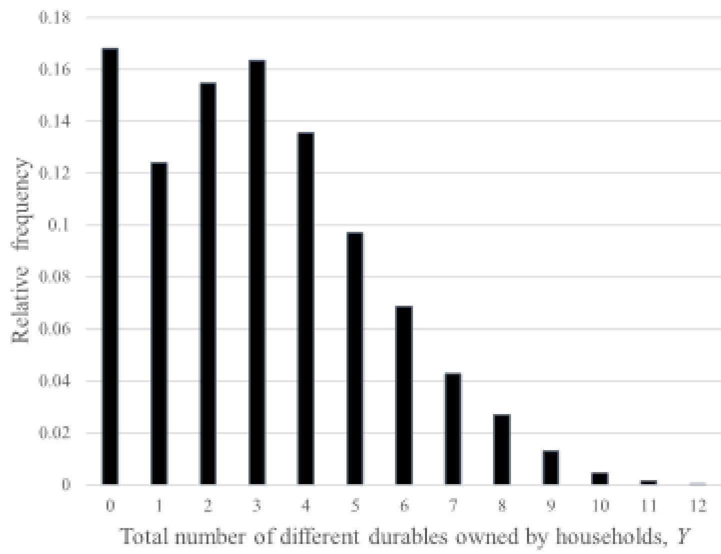
Note that the weight of different durable goods in the sum Y is the same regardless of the price of the good. This may appear to make Y an inadequate measure of affluence, except that there is a "natural weighting" involved in the definition of Y as a simple sum. An expensive item (such as an air conditioner) is more likely to occur in households with more total goods (i.e. higher value of Y) than a cheap item (such as a fan). So, even though it appears that a household with an air conditioner is treated the same as a household with a fan, this is only true when other factors are the same. We surmise that other factors are not the same in a household that has a fan and one that has an air conditioner, and that these dissimilarities will naturally place the ownership of more expensive durable items (air condioners) in households with greater affluence (higher Y), even without using an explicitly higher weight on these premium goods The same argument holds also for household owning different qualities of the same good (e.g. black-and-white vs. plasma tvs) as well as for households owning different numbers of units of the same durable good.

Figure 2 represents the distribution of Y in the sample of households in the urban round of NSS, 1999-00. It represents a draw from the distribution of Y in the urban population, viz. $f(y)$ in (1). What does the shape of the distribution tell us about the distribution of affluence in urban India?

[NSERT FIGURE 2]

approach allows us to determine the *distribution* of households in any class over the relevant expenditures-range, instead of assuming that every household in this range belongs to this class with certainty, as in the cutoffs-based approach (Section 5.4 provides a further discussion of this point).

Fig. 2: Distribution of Y in Urban NSS, 1999-00[*]



[*] See Table A1(a) in the appendix for the tabular presentation of the data herein.

The shape of the distribution in Figure 2 reveals more than one mode, which we saw in Figure 1, might be an indication that it is a mixture of component sub-population densities. But bimodality in itself is not sufficient to justify the existence of sub-populations. The concept of "class" is, however, more than a purely statistical phenomenon – it appears in historical as well as recent literature (see Brittanica), as individual groups in the overall population that have distinct characteristics, as opposed to labels assigned to those with different levels of income or consumption (recall the Pickering-Platt debate, Section 2). In addition, Maitra (2021) shows that when households signal their social status in the marriage market with their observable durable ownership, then total durable expenditure levels of households segregate into clusters, interpretable as "classes". There is, therefore, historical evidence as well as an explicit economic modeling of the class phenomenon, suggesting that a mixture model with classes – as sub-populations – may indeed be an appropriate way to model the durables' ownership distribution in Figure 2. We will call the sub-populations "consumption classes" to underline the fact they are identified using a measure of consumption, viz. durable ownership.

This brings to us to the next question of *how many* classes could generate a distribution such as in Figure 2. Should we look for 3 classes -- a lower, a middle or upper -- that describe the density in Figure 2? Or more, given that there may be a lower middle class and an upper middle class? Note that in an extreme (and completely uninformative) scenario, we could assign each household in our sample (say, of size N) to its own class and achieve a perfect fit to the density in Figure 2 using this $N-$component mixture model! To counter this phenomenon, we impose the criterion that the optimal number of classes is the *smallest* number of sub-populations that yields a good fit to the marginal density in Figure 2. In other words, we iteratively estimate $t-$component mixture models, $t = 2, 3, ...$ The smallest value of t to provide a good fit to the marginal distribution (Figure 2) is the appropriate number of classes that must constitute the urban population in 1999-00.

In Section 5.1, we show that the urban data from NSS 1999-00 is best described by 3 classes, which we will call a lower, a middle and an upper class. In the next section, therefore, we will present the three-component mixture model used to identify the classes and estimate their size and characteristics.

4 A Three-Component Mixture Model to Identify Consumption Classes

Consider 12 durable goods and let Y represent the total number of these goods that a household owns at the time of interview, $Y = 0, 1, 2, ..., 12$. Households can belong to one of three classes – $1, 2$ or 3 – which are defined by the pattern of durables' ownership of members. Denote the density of Y within class i by $\phi_i(y)$ and the density of Y in the population by $f(y)$.[7] A Three-Component Mixture Model postulates the following:

$$f(y) = \pi_1\phi_1(y) + \pi_2\phi_2(y) + \pi_3\phi_3(y) \qquad (5)$$

where π_1, π_2 and π_3 represent the proportions of class 1, 2 and 3 in the population, respectively ($\pi_1 + \pi_2 + \pi_3 = 1$).

To estimate the mixture model – i.e. to estimate the mixing probabilities π_i and the conditional distributions $\phi_i(Y)$ – using maximum likelihood estimation (see Section 4.2), we make an assumption about the family of (discrete) density functions to which ϕ_i $(i = 1, 2, 3)$ belong. We choose the family of binomial distributions, since binomial densities, depending on parameter values, are flexible enough to represent symmetric as well as skewed densities. This feature of binomials gives us the ability to

[7] Remember that Y is a discrete variable, even though we use the term "density function" (and not "mass function") to describe its distribution throughout the chapter.

capture potentially different shapes of different (class-specific) conditional densities of Y.

In particular, assume that a household owns each good with a fixed probability p_i, which depends on the class i ($= 1, 2$ or 3) to which it belongs. The ordering of the p_i's indicates which i ($= 1, 2$ or 3) corresponds to the lower, the middle and the upper class, respectively, since (by definition) $p_L < p_M < p_U$ (L: lower, M: middle, U: upper). Assume that each good is obtained independently by households. Hence the total number of goods owned by a class$-i$ household follows a binomial distribution with parameters 12 and p_i, i.e. $\phi_i(Y) \sim Binomial(12, p_i)$.

Note that we do not claim that households do indeed acquire each good independently and with the same probability. The assumption of binomial densities is a tool to approximate the shape of the conditional densities $\phi_i(Y)$. Think of the process as one of identifying the shapes of the conditional distributions $\phi_i(Y)$; the parameters p_i denote binomial densities that would successfully mimic these shapes.[8]

If the population is a (binomial) mixture of three classes, what is the probability of drawing a household in the sample with a certain level of ownership, say x? This probability – that of observing level x of durable ownership in a household in the sample – is given by:

$$Pr(Y = x; \pi_1, \pi_2, p_1, p_2, p_3) = \pi_1\phi_1(x; p_1) + \pi_2\phi_2(x; p_2) + \pi_3\phi_3(x; p_3) \tag{6}$$

where π_i represents the probability that the household we have drawn belongs to class i and $\phi_i(x; p_i)$ represents the (binomial) probability that a class$-i$ household owns x durables.

The expression in (6) can be used to derive the likelihood function of drawing a sample of size N with durable ownership levels $(y_1, y_2, ..., y_N)$. The likelihood function may then be maximized to obtain maximum likelihood estimates of the parameters, viz. those values of the parameters that would be most likely to have generated the sample we have drawn, viz. $(y_1, y_2, ..., y_N)$. However, maximum likelihood estimation can provide challenges in terms of parameter estimation and hypothesis testing for mixture models (McLachlan and Krishnan (1996)). Calculating likelihoods for a sample based on a mixture model is complicated, and traditional numerical likelihood optimization techniques such as Newton-Raphson break down. Here, I will use the Expectations Maximization (EM) algorithm for likelihood maximization (McLachlan and Krishnan (1996); Dempster et al (1977); Hastie et al (2001)). The EM optimum coincides with the likelihood optimum but is reached (somewhat slowly)

[8] See footnote 9 on the drawbacks of allowing dependence in ownership of each good, or a different probability p_{ij} of ownership for each good i if in class j.

using iterated steps. The algorithm and its application to this analysis are described in Section 4.3.

4.1 Identifiability and Observational Equivalence

Before attempting to estimate the binomial mixture model in $(5)-(6)$, it is necessary to establish that the model is identifiable. In general, a parametric family of densities $f(y;\Psi)$ is identifiable if distinct values of the parameter Ψ determine distinct members of the family of densities $\{f(y;\Psi);\Psi\in\Omega\}$, where Ω is the specified parameter space (McLachlan and Peel (2000)). In other words, a parametric family of densities $f(y;\Psi)$ is identifiable when $f(y;\Psi)=f(y;\Psi^*) \iff \Psi=\Psi^*$.

Blischke (1964) shows that a necessary and sufficient condition for identifiability of binomial mixtures is $h\geq(2r-1)$, where h is the binomial parameter denoting the number of trials and r is the number of components in the mixture. In the current application, $h=12$ (the number of durables) and $r=3$ (the number of classes), so the condition for identifiability is easily satisfied. Hence the model (5) is identifiable.[9]

Note also the issue of observational equivalence known to plague mixture models in general. This means that even when the model is identified as defined above, there is observationally no difference between, for excample, the parameter vector $(\pi_1,\pi_2,1-\pi_1-\pi_2,p_1,p_2,p_3)$ and the vector $(\pi_2,\pi_1,1-\pi_1-\pi_2,p_2,p_1,p_3)$. Observational equivalence makes it hard to uniquely map parameters to class (in the example above: is class 1 the lower class or class 2?). However, the very nature of the current application – the identification of a lower, a middle and an upper class – provides a natural remedy for the issue, since, obviously, $p_L<p_M<p_U$ (L : lower class, M : middle class, U: upper class). Therefore, the ordering of the p_i-estimates tells us which class is the lower class, which is the middle class and which, the upper class.[10]

4.2 Estimation: Maximum Likelihood and the Expectations-Maximization (EM) Algorithm

Having established identifiability, we now proceed to estimation of the mixture model. From (6), we know the probability of observing a household with x durables in the sample. Recall that we have an independent and identically distributed sample of N households where household j is observed to own

[9] In fact, a binomial mixture of up to 6 classes will be identifiable in our model with 12 durable goods, since $h\geq(2r-1)$ is true for $h=12$ and $r\leq 6$.

[10] Recall that if we had assumed different probabilities of ownership p_{ki} for each durable good k ($k=1,2,...,12$) given class i, we would be trying to estimate a vector of 12 probabilities for *each* class! The criterion for mapping $12-$tuple vectors of probability estimates to the lower, middle or upper class would not then be as straightforward as above.

y_j durables ($j = 1, 2, ..., 12$). What is the likelihood of having drawn this sample of households with durables $(y_1, y_2, ..., y_N)$? The likelihood is given by multiplying (given independence) the probabilities in (6), i.e.

$$L(y_1, y_2, ..., y_N; \pi_1, \pi_2, p_1, p_2, p_3) = \prod_{j=1}^{N} [\pi_1 \phi_1(y_j; p_1) + \pi_2 \phi_2(y_j; p_2) + (1 - \pi_1 - \pi_2)\phi_3(y_j; p_3)] \quad (7)$$

where subscript j denotes household j ($j = 1, 2, ..., N$). The log likelihood function is then given by:

$$\text{Log } L(y_1, y_2, ..., y_N; \pi_1, \pi_2, p_1, p_2, p_3) = \sum_{j=1}^{N} [\pi_1 \phi_1(y_j; p_1) + \pi_2 \phi_2(y_j; p_2) + (1 - \pi_1 - \pi_2)\phi_3(y_j; p_3)] \quad (8)$$

The goal of maximum likelihood estimation is to find the values of the parameters $(\pi_1, \pi_2, p_1, p_2, p_3)$ that would maximize the likelihood (7) (hence, the log likelihood (8)) of having drawn the sample we have, viz. $(y_1, y_2, ..., y_N)$. This involves solving first and second order conditions from (8) (obtained by taking partial derivatives with respect to the parameters and setting them equal to zero). It is hard, however, to obtain closed-form expressions for maximum likelihood estimates of the parameters from (7) or (8). The Expectations-Maximization (EM) algorithm is a tool used to simplify difficult maximum likelihood problems such as the above (McLachlan and Krishnan (1996); Dempster et al (1977); Hastie et al (2001)) and is described in Section 4.3. The importance of the EM algorithm lies in its ability to find a path to the maximum likelihood point estimates where traditional numerical techniques typically fail.

4.3 Implementation of the EM Algorithm

Suppose that each household belongs to a particular class and let the dummy variables (δ_1, δ_2) represent the class membership of households, i.e. $\delta_1 j = 1$ if household j belongs to class 1 (0, otherwise), and $\delta_{2j} = 1$ if household j belongs to class 2 (0, otherwise).

Clearly, (δ_1, δ_2) are latent variables since the class memberships of households are unknown. Suppose, however, that (δ_1, δ_2) are known. Then the likelihood and log likelihood functions could be written as:

$$L_{EM}(y_1, y_2, ..., y_N; \pi_1, \pi_2, p_1, p_2, p_3) = \prod_{j=1}^{N} \{\pi_1 \phi_1(y_j; p_1)\}^{\delta_{1j}} \{\pi_2 \phi_2(y_j; p_2)\}^{\delta_{2j}} \{(1-\pi_1-\pi_2)\phi_3(y_j; p_3)\}^{(1-\delta_{1j}-\delta_{2j})} \tag{9}$$

and

$$\begin{aligned} \text{Log } L_{EM}(y_1, y_2, ..., y_N; \pi_1, \pi_2, p_1, p_2, p_3) = \sum_{j=1}^{N} [&\delta_{1j} log\{\pi_1 \phi_1(y_j; p_1)\} + \delta_{2j} log\{\pi_2 \phi_2(y_j; p_2)\} \\ &+ (1 - \delta_{1j} - \delta_{2j}) log\{(1 - \pi_1 - \pi_2)\phi_3(y_j; p_3)\}] \end{aligned} \tag{10}$$

If class memberships (δ_1, δ_2) were indeed known, it would be easy to find close-form expressions for maximum likelihood parameter estimates that maximize Log L_{EM} in (10) above. Since class memberships are not known, the EM algorithm computes the expected values of (δ_1, δ_2) conditional on the data (call these (γ_1, γ_2)), plugs these into (10) and computes the maximands. The procedure is iterated till convergence is obtained. The steps involved are outlined below (McLachlan and Krishnan (1996); Dempster et al (1977); Hastie et al (2001)).

The EM Algorithm for a Three − Component Mixture Model

1. **Start with initial guesses for the parameters**: $(\pi_1^{(0)}, \pi_2^{(0)}, p_1^{(0)}, p_2^{(0)}, p_3^{(0)})$

2. **Expectation (E) step**: at the k^{th} step, compute, as follows, the expected values $(\gamma_i^{(k)})$ of class membership conditional on the data $(y_1, y_2, ..., y_N)$. Note that since class memberships are binary, a household's expected value of class membership conditional on the data is also the estimated probability that a household belongs to class i conditional on the data. That is,

$$\begin{aligned} E(\delta_{ij}/(y_1, ..., y_N; \pi^{(k-1)}, p^{(k-1)}) &= Pr(\delta_{ij} = 1/Y = y_j; \pi^{(k-1)}, p^{(k-1)}) \\ &= \frac{Pr(\delta_{ij} = 1, Y = y_j/\pi^{(k-1)}, p^{(k-1)})}{Pr(Y = y_j/\pi^{(k-1)}, p^{(k-1)})} \\ &= \frac{\pi_i^{(k-1)} \phi_i(y_j, p_i^{(k-1)})}{\pi_1^{(k-1)} \phi_1(y_j, p_1^{(k-1)}) + \pi_2^{(k-1)} \phi_2(y_j, p_2^{(k-1)}) + (1 - \pi_1^{(k-1)} - \pi_2^{(k-1)}) \phi_3(y_j, p_3^{(k-1)})} \end{aligned} \tag{11}$$

Therefore, the conditional expected values ($\gamma_i^{(k)}$) are given by the last term in (11) above, viz.,:

$$\gamma_{ij}^{(k)}$$

$$= E(\delta_{ij}/(y_1, ..., y_N; \pi_1^{(k-1)}, \pi_2^{(k-1)}), p_1^{(k-1)}, p_2^{(k-1)}, p_3^{(k-1)}) \tag{12}$$

$$= \frac{\pi_i^{(k-1)}\phi_i(y_j, p_i^{(k-1)})}{\pi_1^{(k-1)}\phi_1(y_j, p_1^{(k-1)}) + \pi_2^{(k-1)}\phi_2(y_j, p_2^{(k-1)}) + (1 - \pi_1^{(k-1)} - \pi_2^{(k-1)})\phi_3(y_j, p_3^{(k-1)})}$$

$i = 1, 2, 3$; $j = 1, 2, ..., N$.[11]

3. **Maximization (M) step**: at the k^{th} step, compute the parameters as follows. These are the maximands of the EM-log-likelihood function in (10), when (δ_1, δ_2) are replaced by $(\gamma_1^{(k)}, \gamma_2^{(k)})$, viz. the expected values of class membership conditional on the data (calculated in the E step).[12]

$$\pi_i^{(k)} = \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij}^{(k)} \tag{13}$$

$$p_i^{(k)} = \frac{1}{12}[\frac{\sum_{j=1}^{N}\gamma_j^{(k)}x_j}{\sum_{j=1}^{N}\gamma_j^{(k)}} \tag{14}$$

$i = 1, 2, 3$.

4. **Iterate steps 2 and 3 (the E and M steps)** till convergence is obtained in the estimates in (13) $-$(14).[13]

As output, the EM algorithm yields the following estimates:[14]

1. $\hat{\pi}_i$: estimates of the (unconditional) probability that any household belongs to class $i; i = 1, 2, 3$
2. $\hat{p}_i$: estimates of the probability with which a class$-i$ household owns a durable good; $i = 1, 2, 3$

[11] The expression for the conditional probability in (11) $-$ (12) follows from Bayes' Law (see equation (4)).

[12] The expressions in (13) and (14) follow directly from the first order conditions of maximization of (10), with (δ_1, δ_2) replaced by $(\gamma_1^{(k)}, \gamma_2^{(k)})$.

[13] It is straightforward to write a computer program that iterates through the steps of the EM algorithm. The author's code written in STATA is available upon request.

[14] Standard errors of estimates are obtained by taking the square root of the diagonal elements of the (5×5) variance-covariance matrix $[I(\hat{\pi}_1, \hat{\pi}_2, \hat{p}_1, \hat{p}_2, \hat{p}_3)]^{-1}$. Denote the set of parameters by Ψ, viz. $\Psi = (\pi_1, \pi_2, p_1, p_2, p_3)$ and EM-generated estimates $\widehat{\Psi} = (\hat{\pi}_1, \hat{\pi}_2, \hat{p}_1, \hat{p}_2, \hat{p}_3)$. Then, the (5×5) information matrix is given by $I(\widehat{\Psi}) = -\frac{\partial^2 LogL_{EM}(\widehat{\Psi})}{\partial \Psi^2}$; the variance-covariance matrix is given by $[I(\widehat{\Psi})]^{-1}$.

3. $\hat{\gamma}_{ij}$: estimates of the (conditional) probability with which household j (with durable ownership y_j) belongs to class i; $i = 1, 2, 3;\ \ j = 1, 2, ..., N$.

The ownership probabilities $\hat{p}_i$ and the corresponding class-specific (binomial) densities $\phi_i(y, \hat{p}_i)$ answer our motivating question – who are the lower, middle and upper class? – by identifying the distinct ownership patterns (densities) of the different classes. Moreover, the estimates of the unconditional probabilities $\hat{\pi}_i$ – interpretable as estimates of class shares – tells us the sizes of the urban lower, middle and upper classes in 1999-00. Finally, the estimated (conditional) probabilities of class membership, $\hat{\gamma}_{ij}$, along with $\hat{\pi}_i$ and $\hat{p}_i$, enable a random assignment of each household into a particular class. This allows a descriptive analysis of other class-specific household characteristics such as per capita monthly expenditures, education of household heads, household types by employment and so on (see Section 5.4). The next section presents the results.

5 Results and Discussion

5.1 Estimates

The estimates produced by the EM algorithm are presented in Table 1. The numbers in column (1) of Table 1 represent the population share of each class, $\hat{\pi}_i$. The middle class is estimated to constitute 62% of urban households. This is roughly equivalent to 17% of the total population, given that urban households accounted for about 28% of all Indian households in 2001 (2001 census, Indiastat, http://www.indiastat.com). The lower and upper classes are found to constitute 20 and 18% of urban households, respectively. Asymptotic standard errors (obtained from the information matrix) are small, supporting the existence of three classes in the population.

[INSERT TABLE 1]

Table 1: Estimates from a Three-Component Mixture Model using Durable Ownership Y

Urban Sub-Sample, Indian NSS, 55th Round (1999-00), N = 48,924 households

Sub-population (Class)	Mixture Estimate (Std. Error)		
	(1) Share of Urban Population $\widehat{\pi}_i$	(2) Probability of Owning a Good $\widehat{p}_i$	(3) Mean No. of Goods Owned (of 12)* $12\widehat{p}_i$
Lower (L)	0.2034 (0.005)	0.0257 (0.002)	0.3084 (0.007)
Middle (M)	0.6161 (0.005)	0.251 (0.003)	3.012 (0.01)
Upper (U)	0.1804 (0.006)	0.5249 (0.004)	6.2988 (0.014)

* The 12 goods include 5 recreational goods (record player, radio, tv, vcr/vcp, tape/cd player), 4 household goods (electric fan, a/c, washer, fridge) and 3 trasnport goods (bicycle, motor bike/scooter, motor car/ jeep)

Column (2) reports estimates of the probability parameter p_i for each class $i = L, M, U$. Lower class households are found to own a good with 3% probability while middle and upper class households own a good with probabilities of 25% and 52% respectively. Small standard errors support three distinct patterns of durables consumption behaviour.[15]

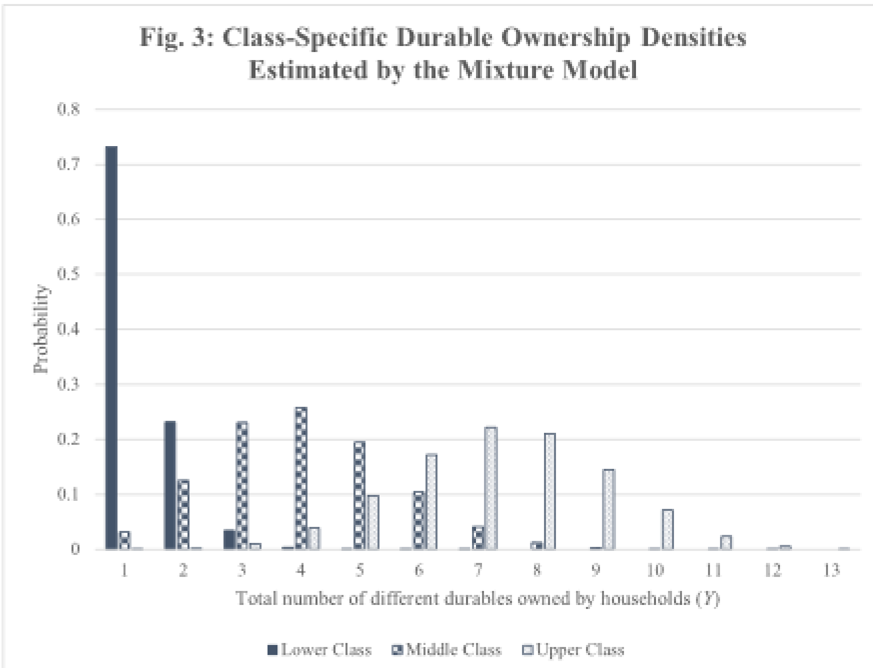
The mean number of durable goods (out of 12) owned by class$-i$ households is simply $12p_i$ (the mean of the binomial distribution for class i). These estimates are reported in Column (3) of Table 1. The lower, middle and upper classes are found to own, on average, 0.3, 3 and 6.3 goods, respectively.

Figure 3 plots the binomial density functions ϕ_i at the estimated parameters $\hat{p}_i$ $(i = 1, 2, 3)$. It is the shape and structure of each density $\phi_i(y; \hat{p}_i)$ that identifies each class i in our analysis. The ownership density of the lower class and upper classes (binomials with probability parameters 0.03 and 0.52, respectively) are slightly positively skewed whereas that of the middle class is symmetric. Recall that we used the family of binomial functions in our mixture model for precisely this purpose – to allow for the estimation of different shapes of the different ϕ_i functions depending on the estimated

[15]The estimates (standard errors) of the differences are as follows: $\hat{p}_L - \hat{p}_U = -0.5$ $(0.004), \hat{p}_L - \hat{p}_M = -0.23$ (0.002) and$\hat{p}_U - \hat{p}_M = 0.27(0.003)$ ($L \approx$Lower; $M \approx$Middle; $U \approx$Upper).

parameters $\hat{p}_i$. Recall also that in estimating the densities ϕ_i, we made no explicit assumptions about *who* constitutes the classes other than that the three classes are different and that higher classes are more affluent than lower classes. The criteria for class membership – viz. the location and shape of the densities ϕ_i – were instead delivered by the estimation process, based on natural clusters in the data.

[INSERT FIGURE 3]



Finally, note the overlapping sections of the density functions of the different classes, which indicate that the level of durables owned do not uniquely identify classes (as in cutoffs-based approaches). In Section 5.3, we will specifically calculate the probabilities with which a household with durables x belongs to each of the classes.

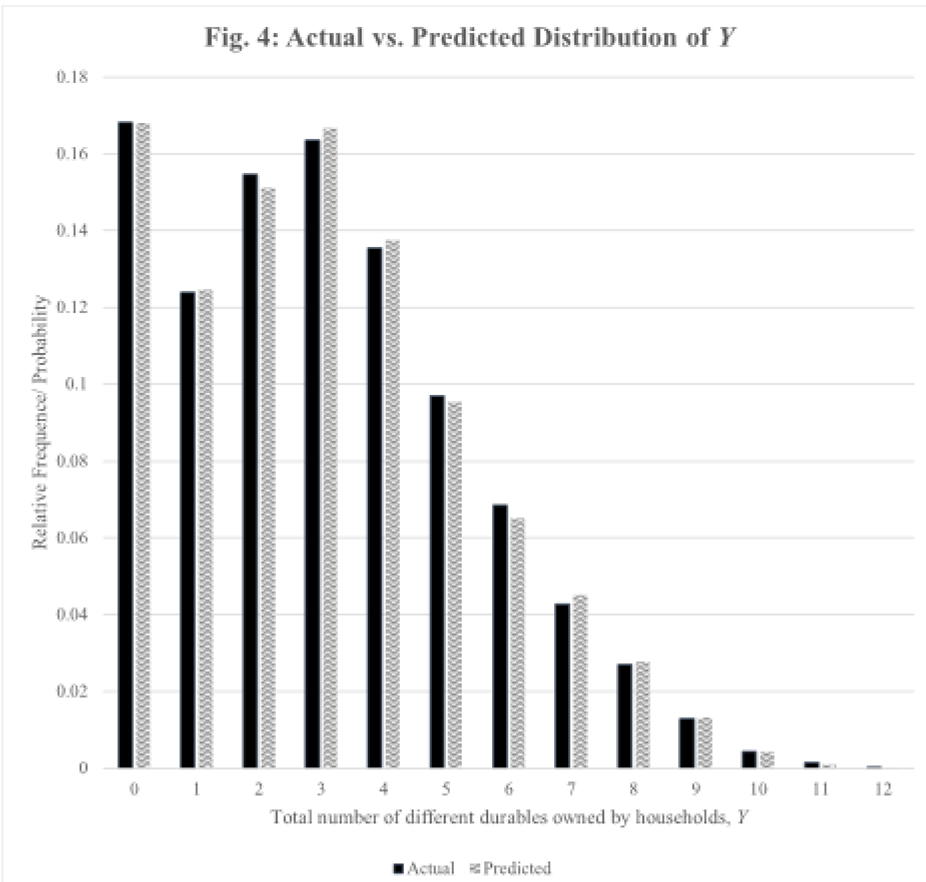
Let us now construct the ***predicted*** marginal density of durable ownership from our estimated parameters. If the population constitutes 3 classes in the proportions $(\hat{\pi}_1, \hat{\pi}_2, 1 - \hat{\pi}_1 - \hat{\pi}_2)$ and with densities $\phi_1(y; \hat{p}_1)$, $\phi_2(y; \hat{p}_2)$ and $\phi_3(y; \hat{p}_3)$, respectively, then what is the predicted probability of observing any level of durable ownership, say y^*, in this population? This is simply the expression in (6), with parameter values set equal to the estimates. That is,

$$\hat{Pr}(Y = y^*; \hat{\pi}_1, \hat{\pi}_2, \hat{p}_1, \hat{p}_2, \hat{p}_3) = \hat{\pi}_1 \phi_1(y^*; \hat{p}_1) + \hat{\pi}_2 \phi_2(y^*; \hat{p}_2) + \hat{\pi}_3 \phi_3(y^*; \hat{p}_3) \tag{15}$$

$y^* = 0, 1, 2, ..., 12.$

Having calculated the predicted marginal density of Y as in (15), we can now answer the question: how well does the predicted marginal density match the observed marginal density of Y? Figure 4 plots the actual relative frequency of (Y) observations in the data along with the predicted values. The figure demonstrates a very good fit to the data.

[INSERT FIGURE 4]



Recall, from Section 3, our assertion that the number of classes that describe the data must be the smallest number of sub-populations that yields a good fit to the observed marginal density. Consider then the findings from a Two-Component (two classes) Mixture Model fitted to the data by EM plotted in Figure 5. The fit is visibly worse than that of the Three-Component Model. Hence, 3 appears to be the minimum number of classes that provides a good fit to the data.

[INSERT FIGURE 5]

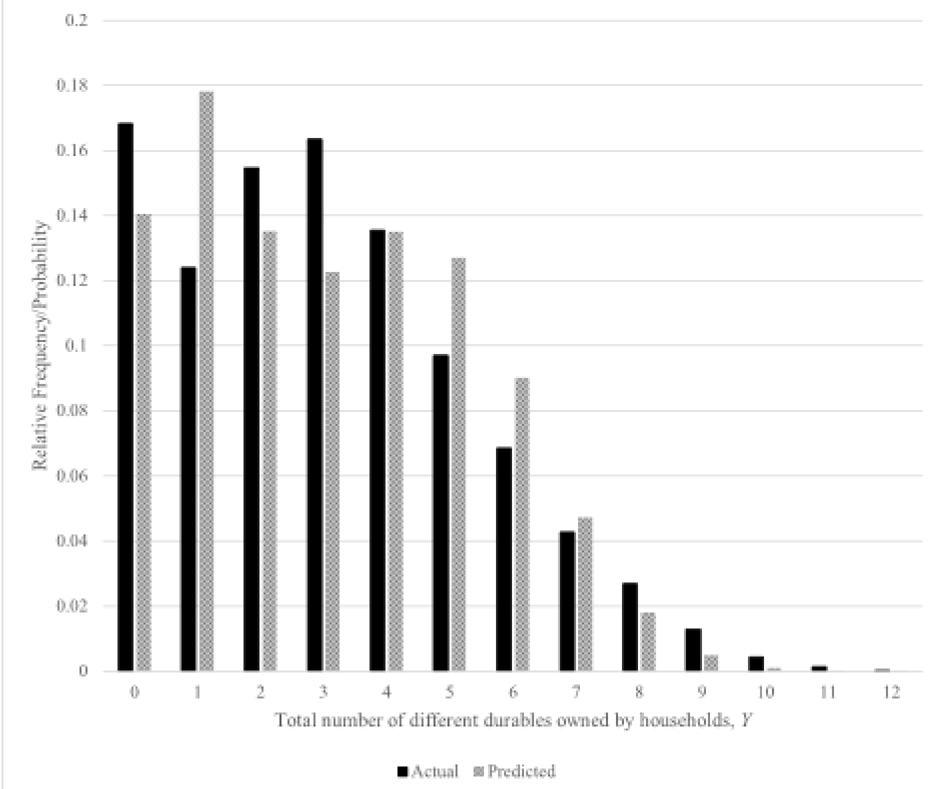


Fig. 5: Two-Component Mixture Model*:
Actual vs. Predicted Distribution of Y

* Parameter Estimates, Two-Component (Binomial) Mixture Model, Urban NSS, 1999-00

Class	$\widehat{\pi}_i$	$\widehat{p}_i$	$12\widehat{p}_i$
Lower	0.4	0.1	1.08
Upper	0.6	0.4	4.52

5.2 What happened to urban consumption classes in India, post-liberalization?

The ultimate goal of our analysis was to examine the fate of urban consumption classes in India – their sizes and characteristics – after the liberalization policies of 1991 came into effect. Maitra (2016, 2017) look at various aspects of this comparison using the mixture analysis outlined above, using the NSS rounds of 1993-94, 1999-00 and 2004-05. The use of durable ownership data – immune to known measurement issues in the NSS, 1999-00 round – to define the classes allows us to make credible comparisons across these rounds of data. Moreover, the mixture approach precludes the need to make arbitrary assumptions about who constitutes the classes in each survey year. These class-definitions are, in fact, delivered by the mixture approach, which uses natural clusters in the data in each round to identify class-specific conditional ownership densities $\phi_i(\widehat{p}_i)$ in each survey year.

Table 2 presents Maitra's (2017) mixture results for urban classes in the NSS rounds of 1993-94, 1999-00 and 2004-05.[16]

[INSERT TABLE 2]

Table 2*: Estimates from Three-Component Mixture Models, Urban NSS subsamples, 1993-94, 1999-00 & 2004-05

Year	π_L	π_M	π_U	p_L	p_M	p_U
1993-94	0.324	0.647	0.029	0.085	0.313	0.644
1999-00	0.200	0.621	0.179	0.035	0.341	0.590
2004-05	0.161	0.603	0.235	0.079	0.340	0.627

* This table is reproduced from Maitra (2017).

Notice the contraction in the size of the lower and middle classes over the decade 1993-94 to 2004-05, while the upper class grows larger over this time. In terms of class definitions, the probability of durable ownership (p_L) of the lower class dropped dramatically between 1993-94 and 1999-00 but seems to have caught up again in 2004-05 to the original 1993-94 level. The upper class seems to have stagnated a bit over this time (p_U has not increased), while the middle class seems to be doing somewhat better over the period considered. Taken together, the estimates of class size and probabilites suggest a significant wave of upward class mobility among urban households in India in the 1990s.[17]

5.3 Mixture-estimated Probabilities of Class Membership ($\hat{\gamma}_i(y)$)

Let us return to our baseline mixture estimates for 1999-00 in Table 1, where 12 durable goods are used in the definition of Y. The overlapping sections of the class-specific densities suggest that certain levels of durable ownership may be compatible with households in more than one class. Can we calculate the specific probability $\hat{\gamma}_i(y)$ that a household belongs to different classes i $(= 1, 2, 3)$ conditional on the number of durables y owned?[18] Using Bayes' Law, as in equation (4), we can derive these specific probabilities to be:

[16]The analysis presented in Table 2 uses a total of 8 (of the current 12) durable goods since these were the common items asked about in the questionnaires across all the survey years. The results for 1999-00 urban classes are very similar whether 8 or 12 durables are used. Also, 3 urban classes were found to be optimal in each of the years examined.

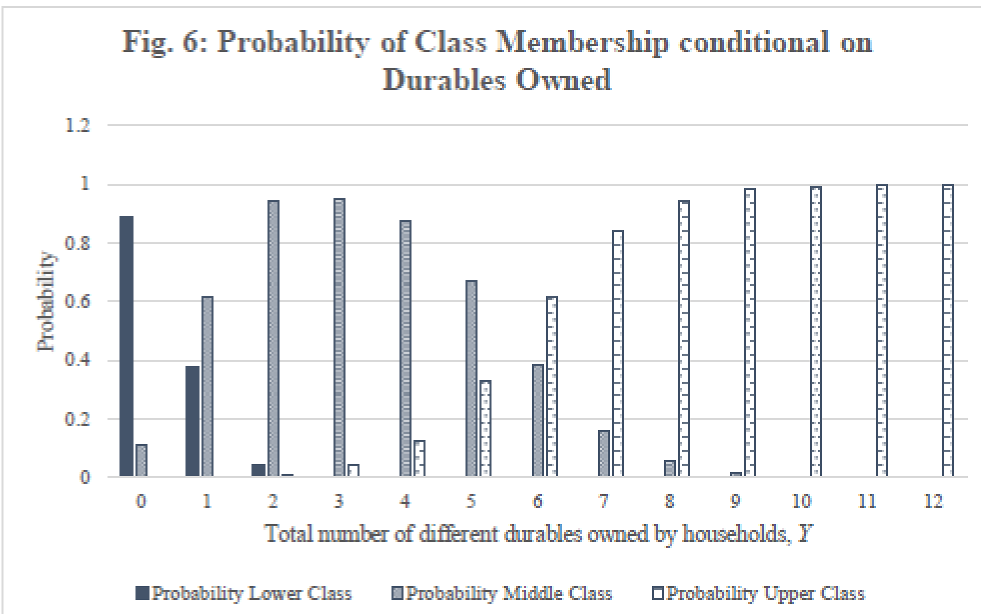
[17]See Maitra (2016, 2017) for a detailed discussion of the implications of these estimates for poverty and inequality in India in the 1990s.

[18]Clearly, $\hat{\gamma}_1(y) + \hat{\gamma}_2(y) + \hat{\gamma}_3(y) = 1$ for any $y = 0, 1, 2, ..., 12$, since any household must belong to one of the 3 classes.

$$\hat{\gamma}_i(y) = \frac{\hat{\pi}_i \phi_i(y, \hat{p}_i)}{\hat{\pi}_1 \phi_1(y, \hat{p}_1) + \hat{\pi}_2 \phi_2(y, \hat{p}_2) + \hat{\pi}_3 \phi_3(y, \hat{p}_3)} \tag{16}$$

Figure 6 plots the probabilities $\hat{\gamma}_i(y)$ that a household belongs to different classes i $(= 1, 2, 3)$ conditional on the number of durables y owned. Evidently, households with low values of Y are most likely to belong to the lower class (class 1) whereas those with the highest values of Y are almost certain to belong to the upper class (class 3). Hence, unlike in previous studies that use the cutoffs-based approach, the current approach places households (arguably, more realistically) in different classes with a probability rather than with certainty.

[INSERT FIGURE 6]



In the next section, we will discuss how to use the estimated probabilities $\hat{\gamma}_i(y)$ to assign each household to a class. The assignment will allow us to look at characteristics (other than Y) of each class, further illuminating our understanding of who constitutes the different classes.

5.4 Class Characteristics: A Descriptive Analysis

We have shown that the mixture approach assigns households to different classes with a probability rather than with certainty. Using these (conditional) probability estimates, it is possible to estimate the number of observations of each value of Y that belongs to each class. Based on this computation, we can randomly assign households to classes. Here is an example of how the assignment is performed.

Suppose we estimate $\hat{\gamma}_1(0) = 0.6$, $\hat{\gamma}_2(0) = 0.1$ and $\hat{\gamma}_3(0) = 0.3$. This means that a household that owns none of the twelve durable goods (i.e. $Y = 0$) belongs to class 1 with probability 0.6, class 2 with probability 0.1 and class 3 with probability 0.3. Now suppose that in the dataset, there are 100 observations for $Y = 0$. We can randomly assign 60 of these 100 households with $Y = 0$ to class 1 (consistent with $\hat{\gamma}_1(0) = 0.6$), 10 to class 2 (consistent with $\hat{\gamma}_2(0) = 0.1$) and 30 to class 3 ($\hat{\gamma}_3(0) = 0.3$). The same procedure could be followed to randomly assign households to classes for each other value of Y (i.e. $Y = 1, 2, .., 12$, respectively).[19]

Assigning a class to each household allows a descriptive analysis of the characteristics of each class. Below, in Tables 3-4 and Figures 7-14, we will look at the durable ownership patterns for specific goods as well as a host of socioeconomic characteristics.

Table 3 and Figure 7 reveal the durables consumption patterns of households belonging to the three classes (assigned by the procedure described above). Recreational and household goods appear to be more commonly owned by all classes than are transport goods.[20] Of these, electric fans and televisions are most popular among the top two classes, whereas fans and bicycles are most popular among the lower class.

[INSERT TABLE 3(a) AND TABLE 3(b)]

[19]But randomly assigning households to classes can lead to the same household being placed in different classes in different rounds of assignments (or draws)! This could result in different class-specific characteristics being obtained in different draws. To address this issue, we could repeat the procedure of random assignments multiple times. We could observe any characteristic, say Z, in the sample after each set m of random assignments ($m = 1, 2, ..., M$) and report the average values of these characteristics across all the draws $[\frac{1}{M}\sum_{m=1}^{M} z_m]$ along with the standard deviations $[\frac{1}{M}\sum_{m=1}^{M}(z_m - \frac{1}{M}\sum_{m=1}^{M} z_m)^2]^{0.5}$. This process of repeated draws (here, rounds of random assignments) from the same sample forms the essence of bootstrapping techniques (Hastie et al (2001)). (The results presented in Tables 3-4 and Figures 7-14 are, however, obtained from a single draw as they serve a descriptive purpose.)

[20]This could be partly attributable to the fact that, among the 12 goods considered, there are more recreational and household goods (5 and 4, respectively) than there are transport goods.

Table 3(a): Ownership by Durable Categories by Class in the Urban Sub-sample, NSS 1999-00, N = 48, 924 households

Category (Class)	Mean No. of Goods Owned by Households				Proportion of Households Owning At Least one Good in the Relevant Category, by Class			
	All (12 items)	Recreation Goods (5 items)	Household Goods (4 items)	Transport Goods (3 items)	All (12 items)	Recreation Goods (5 items)	Household Goods (4 items)	Transport Goods (3 items)
Lower (L)	0.31	0.12	0.11	0.07	0.27	0.12	0.11	0.07
Middle (M)	3.01	1.37	1.06	0.58	0.97	0.85	0.79	0.53
Upper (U)	6.30	2.51	2.52	1.27	1.00	1.00	0.99	0.87

Table 3(b): Ownership of Individual Durable Goods by Class in the Urban Sub-sample, NSS 1999-00, N = 48, 924 households

Proportion of Households Owning the Relevant Good, by Class

Category (Class)	Recreational Goods					Household Goods				Transport Goods		
	Record Player	Radio	TV	VCR/ VCP	Tape/ CD Player	Electric Fan	Air Cond.	Washing Machine	Fridge	Bicycle	Motor Bike/ Scooter	Motor Car/ Jeep
Lower (L)	0.00	0.07	0.04	0.00	0.01	0.11	0.00	0.00	0.00	0.07	0.00	0.00
Middle (M)	0.01	0.39	0.68	0.02	0.27	0.77	0.07	0.04	0.18	0.43	0.14	0.01
Upper (U)	0.05	0.58	0.97	0.19	0.71	0.97	0.41	0.39	0.75	0.53	0.60	0.14

[INSERT FIGURE 7(a) AND 7(b)]

Fig. 7(a): Ownership by Durable Categories by Class, Urban Sub-sample, NSS 1999-00

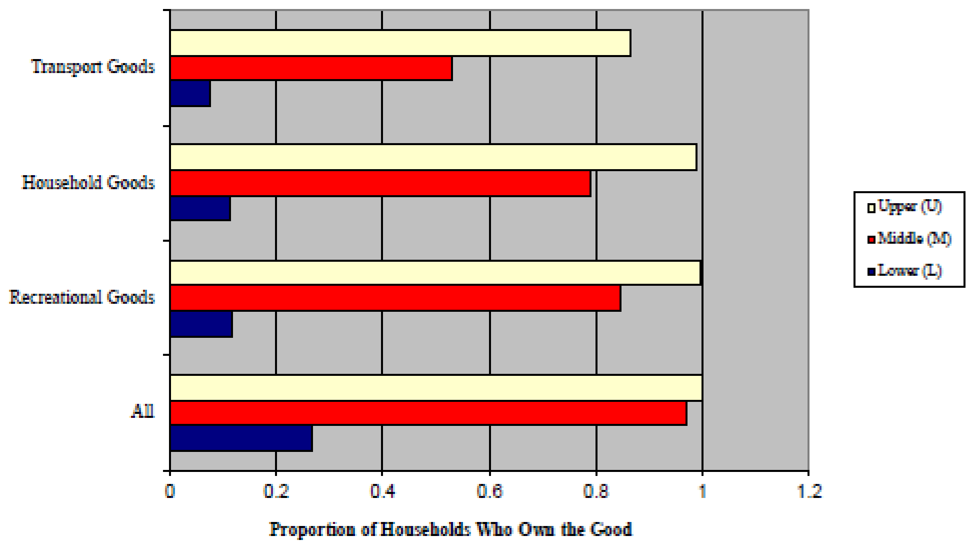


Fig. 7(b): Ownership of Individual Goods by Class, Urban Sub-sample, NSS 1999-00

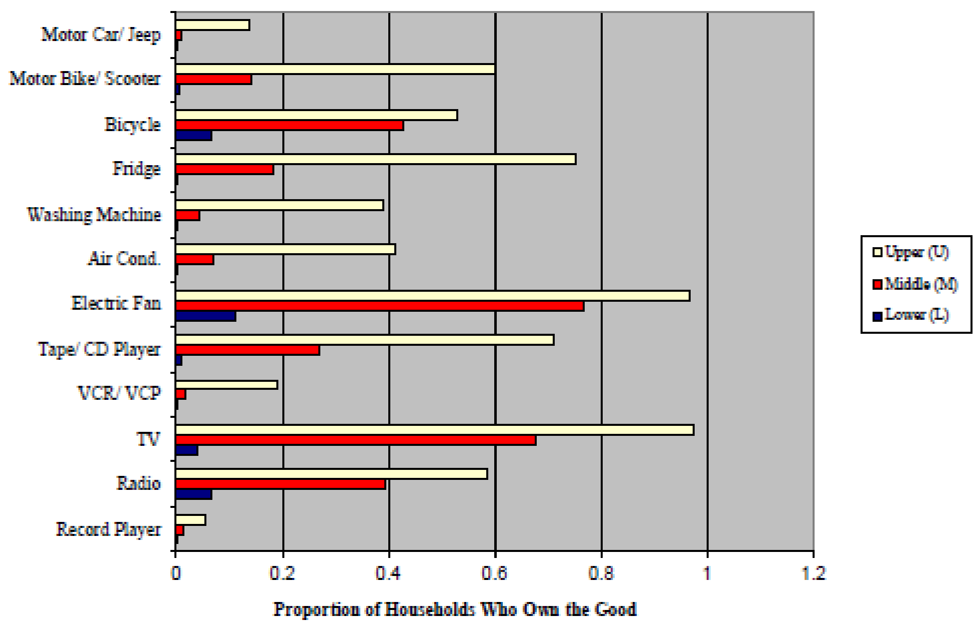


Table 4 reports the per capita monthly expenditures (PCE) of households in each assigned class. Note, yet again, that the fundamental difference between the cutoffs-based approach and the mixture approach lies in the postulated *distribution* of households: the cutoffs-based approach assumes every household with PCE in a certain range to belong to a particular class, whereas the mixture aproach identifies classes whose expenditures are distributed over a PCE-range. Consequently, the ranges of

PCE of different classes are overlapping instead of being mutually exclusive.

[INSERT TABLE 4]

Table 4: Household Characteristics, by Class, in the Urban Sub-sample, NSS, 55th Round (1999-00)

Category (Class)	Per Capita Monthly Household Expenditure in 2000 Rupees [2005 US$, PPP Converted]									Other Household Characteristics		
					Percentiles							
	Mean	Std. Dev.	Min.	Max.	25	50	75	90	99	Avg. No. of Meals Per Day Per Person (Mean)	Proportion of Literate Household Members (Mean)	Household Size (Mean)
Lower (L)	791.26	859.11	17	50528	423	625	981	1421	2791.43	2.34	0.64	3.97
	[65.53]	[71.15]	[1.41]	[4184.83]	[35.03]	[51.76]	[81.25]	[117.69]	[231.19]			
Middle (M)	961.79	1772.39	49	205987	532	762	1140	1663	3485	2.38	0.77	4.65
	[79.66]	[146.79]	[4.06]	[17060.27]	[44.06]	[63.11]	[94.42]	[137.73]	[288.63]			
Upper (U)	1469.57	1109.97	224	35612	842	1229	1777	2490.6	5390.08	2.41	0.88	5.12
	[121.71]	[91.93]	[18.55]	[2949.46]	[69.74]	[101.79]	[147.17]	[206.28]	[446.42]			

Addendum: Percentiles of Per Capita Monthly Expenditure (2000 Rupees) in the Entire Sample, N = 48, 921

Percentile	10	20	30	40	50	60	70	80	90	99
Value	392	490	584	686	801	940	1120	1377	1815	3799.56

Note: CPI inflation from 2000 to 2005: 1.215 (Indian Labour Bureau); PPP conversion rate INR/USD: 14.67 (ICP 2005)

Keeping in mind this difference in approaches, how do the PCE of the mixture-identified classes roughly compare with those assumed by previous studies? Take the middle class, for example. The 1^{st} and 99^{th} percentiles of the daily per capita expenditure of the middle class identified here are about \$0.75 and \$9.62 (at 2005 PPP), with the median being \$2.10 (see Table 4: the addendum). Banerjee and Duflo (2008) define the middle class as having a daily per capita expenditure of \$2−\$4 or \$6−\$10 at 1993 PPP (roughly \$2.68−\$5.36 and \$8.04−\$13.40 at 2005 PPP). Ravallion's (2010) middle class has daily per capita expenditures of \$2−\$13 at 2005 PPP. [21] Also, the mean number of durables owned

[21] The cutoffs defined by Easterly (2001) and Birdsall, Graham and Pettinato (2000) while not defined specifically for developing nations or India would capture only the lower end of middle class PCEs identified here. For instance, by Easterly's (2001) denition of the middle class (those lying between the 20^{th} and 80^{th} percentile of the consumption distribution), the daily PCE cutoffs would be \$1.35 and \$3.80 (2005 PPP-adjusted). Birdsall, Graham and Pettinato's (2000) denition (those lying between 75% and 125% of median income) would yield cutoffs of \$1.66 and \$2.76 (2005 PPP-adjusted). Birdsall's (2010) cutoff of \$10 and above (albeit for a denition of the "indispensable middle class") would

by the middle class as per Banerjee and Duflo's (2008) definition is found to be 3.77, which is very close to the mean durables ownership (3.01) of the middle class identified herein. The middle class as per Sridharan's (2004) definition is found to own on average 5.55 durables, a considerably higher figure.

The differences in distributional assumptions of the mixture versus the cutoffs-based approaches implies, furthermore, that estimates of any class characteristic that is sensitive to distribution could be very different based on which approach is used, even when the range of class-specific expenditures is comparable across approaches (as in Banerjee and Duflo (2008) or Ravallion (2010)). As an example, consider the estimate of the *size* of the middle class as a proportion of urban households. Using Banerjee and Duflo's (2008) definition of middle class in our (NSS, 1999-00) sample, the size of the middle class is obtained to be 32%. Ravallion's definition yields a middle class of 56%. The mixture estimate obtained using the NSS, 1999-00 sample suggests a middle class that comprises 62% of urban households.[22]

Figure 8 plots the education levels of the household head, by class. The lower class has the highest component of illiterate heads (32%) whereas the upper class has the highest component of heads with a graduate degree (38%). Middle class household heads are most likely to have secondary education (18%) although graduates comprise a comparable component as well (15%). A large proportion (18%) of middle class heads appear to be illiterate. This finding appears surprising if the perception of the middle class as being largely white-collar workers is true. However, this finding would be consistent with an environment of active social mobility post-liberalization, characterized by a large influx of lower class members into the middle class.

[INSERT FIGURE 8]

exclude most of the middle class identied herein.

[22]It may not be informative to compare the size-estimates reported (implicitly) in Banerjee and Duflo (2008) and Ravallion (2010) with the mixture estimates obtained herein, since these estimates are derived using data from completely different years. Differences in size-estimates in these versus the current study may not be unambiguously attributed to the difference in approach, since the size of the middle class could be changing over time too. Sridharan's (2004) estimates, on the other hand, are comparable to the mixture estimates in this paper, since the data used therein correspond roughly to the same years (1998-99).

Fig. 8: Education of Household Head, by Class

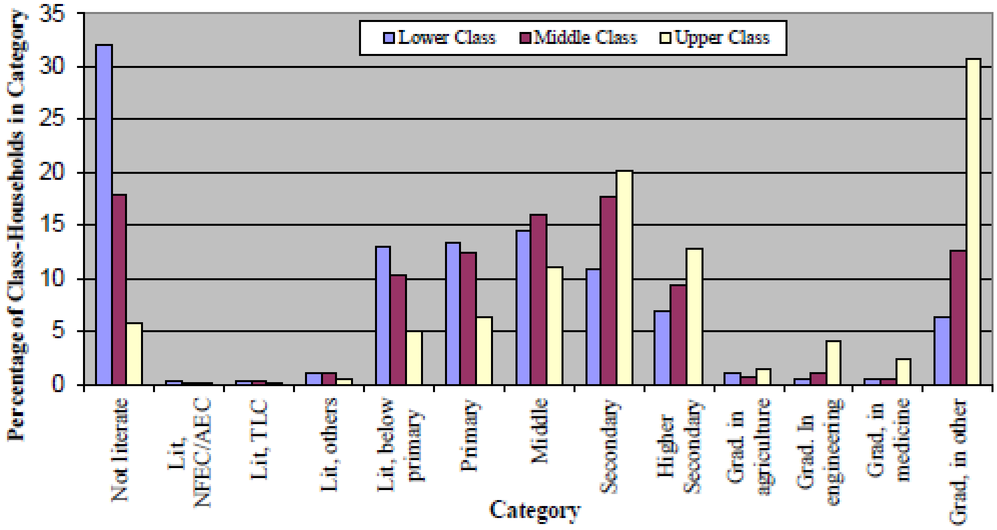
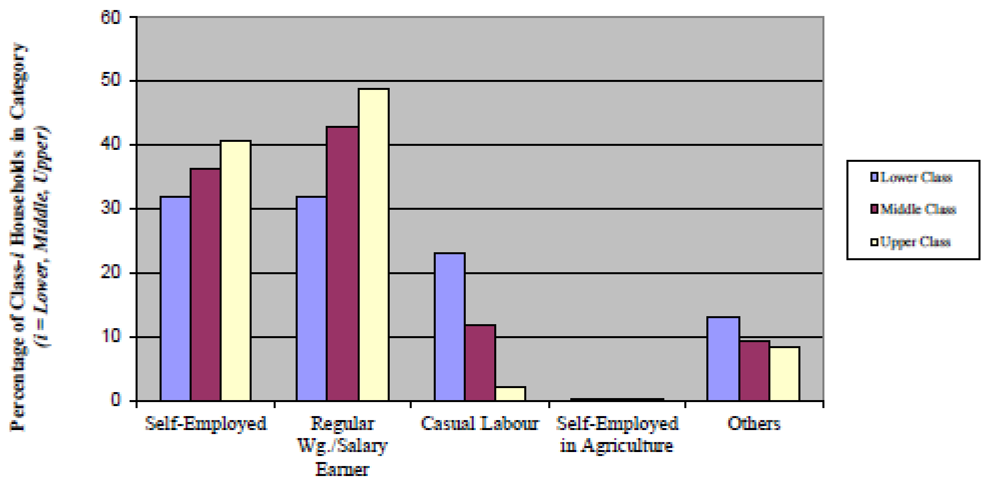


Figure 9 presents a plot of household type by employment. Being an urban sample, the proportion of households who are self-employed in agriculture is negligible. The largest component of households in each class are wage/salary earners. This fact is also mirrored in Figure 10 which plots sources of household income. Over 50% of households in each class have reported income in the past year from wages and salaries. Income from non-agricultural enterprises is reported by more than 30% of households in each class. A large proportion of households also report owning land. Income from interests and dividends is the third most highly reported source of income by the top two classes 15% and 7% of upper and middle class households, respectively. For the lower class, income from other sources is reported by considerably more households (12%) than is income from interests and dividends (2%).

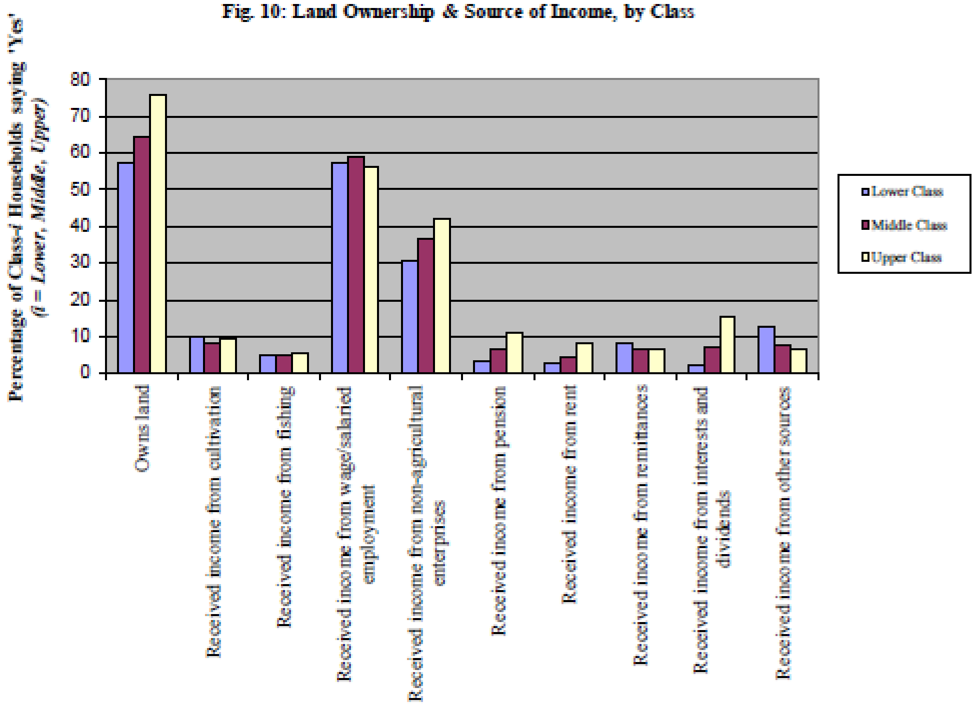
[INSERT FIGURE 9]

Fig. 9: Type of Employment, by Class



[NSERT FIGURE 10]

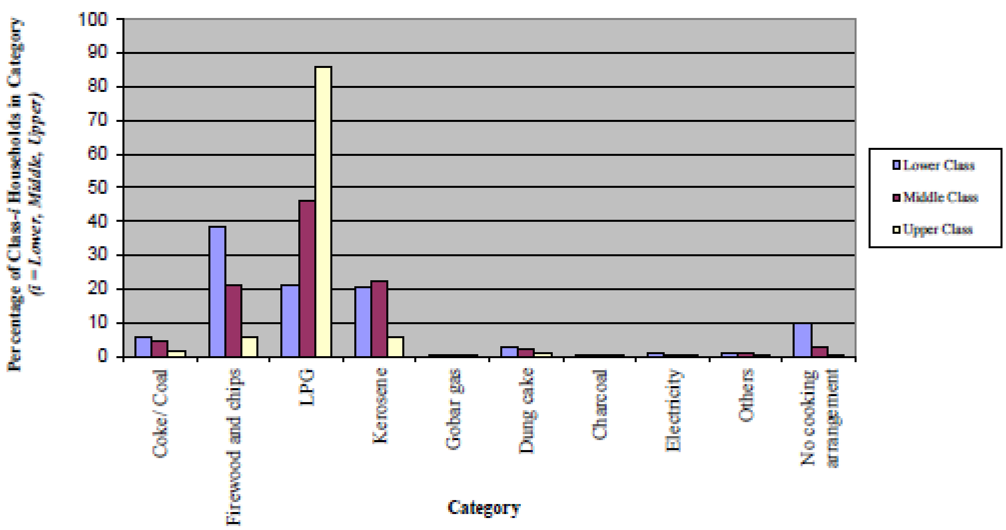
Fig. 10: Land Ownership & Source of Income, by Class



Figures 11 and 12 present a summary of the primary sources of energy used in cooking and lighting. LPG is most commonly used for cooking among the top two classes; firewood and chips are most common among lower class households. For lighting, electricity is most common in all classes, although 25% of lower class households use kerosene as the primary source of energy.

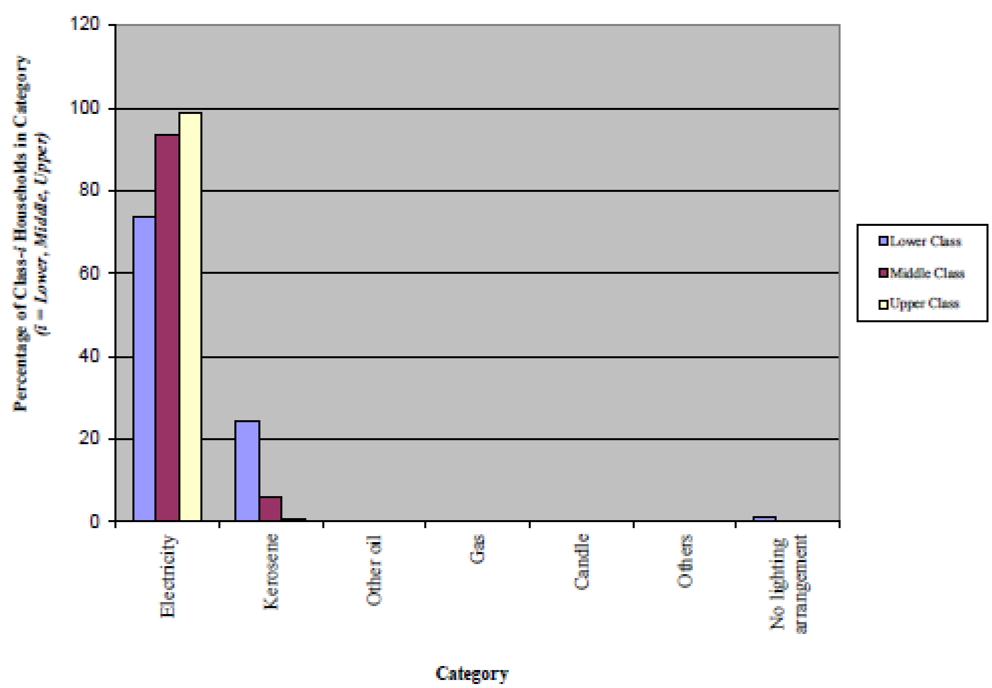
[INSERT FIGURE 11]

Fig. 11: Primary Source of Energy Used for Cooking, by Class



[INSERT FIGURE 12]

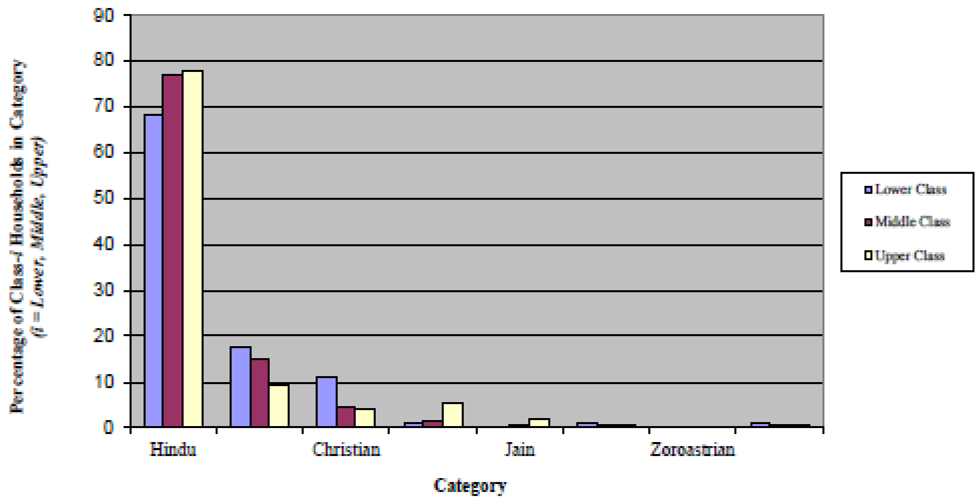
Fig. 12: Primary Source of Energy Used for Lighting, by Class



Finally, Figures 13 and 14 provide a summary of class composition by religion and social class. Hinduism is the religion of the majority in India, so it is not a surprise that Hindus constitute the largest component of all classes. However, Muslims and Christians form a larger component of the lower class (18% and 11%, respectively) than the middle and upper classes (15% and 4% of the middle class while 10% and 4% of the upper class are Muslim and Christian, respectively). Scheduled Castes and Tribes also form a larger component of the lower than the middle and upper classes.

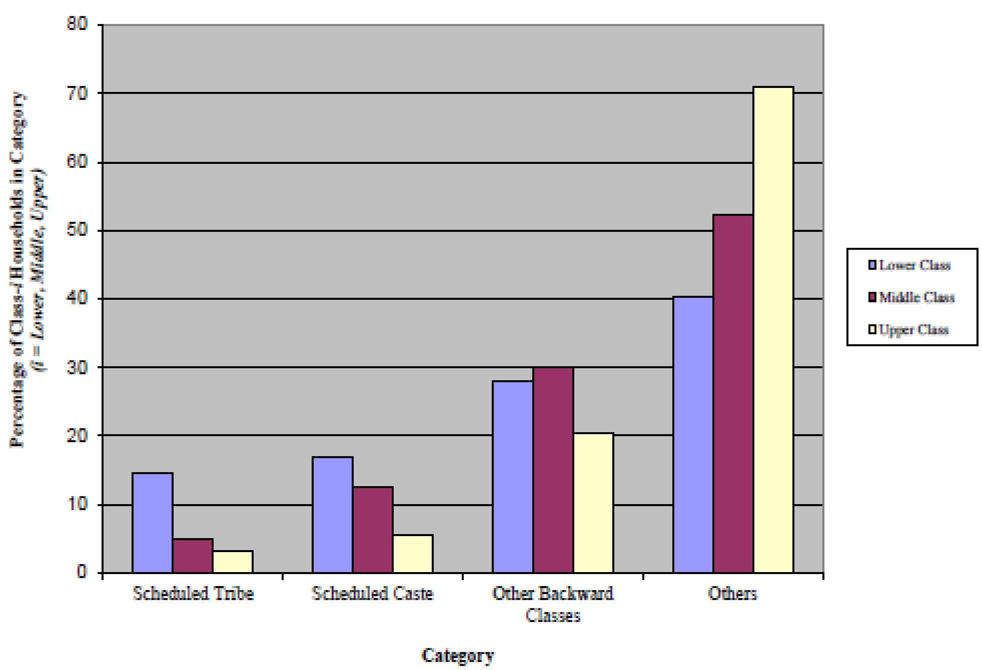
[INSERT FIGURE 13]

Fig. 13: Religion, by Class



[INSERT FIGURE 14]

Fig. 14: Social Group, by Class



The descriptive analysis presented above demonstrates how mixture estimates – computed using durable ownership data – may be used to assign households to classes (albeit probabilistically), which further permits an examination of class-specific characteristics *other* than durable ownership (per capita household expenditure levels, for example). In this sense, our methodology presents a "dual" (or reverse) strategy for class identification as compared with traditional approaches that use expenditure cutoffs, primarily, to identify classes and then infer the durable ownership of classes from the expenditures-based classification.

6 Alternative Approaches to Identifying Classes: $K-$Means Clustering

An alternative approach to examining natural clusters in data on durable ownership would be the use of a traditional clustering model such as the K-Means Model (Hastie et al (2001)). In this method, the iterative algorithm to identify clusters (or classes) attempts to divide data points into partitions, so as to minimize the deviation of all members of a cluster from a central point in the cluster. There is no statistical model involved in the process and the clusters – being partitions – do not intersect. Hence, the K-Means method assigns each datapoint to a class with certainty based on its proximity to a central point in that class.

The goal of a mixture model, on the other hand, is to identify sub-populations with the help of a statistical model (here, binomial mixtures) so as to best fit the observed distribution of the data points. The use of a statistical model allows the use of traditional methods of inference (e.g. calculating standard errors, formulating tests etc). More importantly, the mixture process leads to a probabilistic assignment of data points to sub-populations. The overlapping "fuzzy" class boundaries are especially relevant to our current application as they are interpretable as zones of class "transitions", a natural component of our research question – what happened to consumption classes in India post-liberalization?

7 Conclusion

Mixture models have been examined in depth in the statistical literature since the 1950s. However, they have not been used widely in economic applications due to complexities in interpretation. How do we justify the existence of sub-populations? How do we assign "names" to the sub-populations that are identified when all that the mixture model tells us is that they are different? How can we be sure that the estimates we obtain are unique?

In this chapter, I present and describe the technical aspects of a three-component binomial mixture model, which is then used to identify urban Indian consumption classes based on their total durable ownership. The approach is well-suited for this task as class-definitions emerge from the process – based on natural clusters in the data – instead of having to be specified by researchers. Interested readers are encouraged to consult the larger literature on mixture models and clustering methods (see the section below titled "Recommended Reading") for further discussions of the topic, including mixture analysis of more complex relationships within sub-populations.

An important goal of this chapter has been to provide a conceptual interpretation of the mixture process, so as to demonstrate how it may be used meaningfully despite the complexities of interpretation described above. Failing to understand or address the conceptual background for a mixture model comes with the danger of reducing it to an arbitrary procedure for data fitting with no real informational content. I hope the current demonstration underlines these lessons amply, and that it will inspire more studies that effectively use mixture models to illuminate economic phenomena.

Recommended Reading

- Everitt, B. S., Hand, D. J., 1981. Finite Mixture Distributions Monographs on applied probability and statistics, London; New York: Chapman and Hall.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics.
- Maitra S., 2021. On the Theory and Measurement of Relative Poverty using Durable Ownership Data. Working Paper. https://www.sudeshnamaitra.com/Maitra_relative_poverty_Sep2021.pdf . Accessed 5 January 2022.
- Maitra, S., 2017. Class Matters: Tracking Urban Inequality in Post-Liberalization India using a Durables-based Mixture Model. Applied Economics Letters. 24(17), 1203-1207.
- Maitra, S., 2016. The Poor Get Poorer: Tracking Relative Poverty in India using a Durables-based Mixture Model. Journal of Development Economics. 119, 110-120.

- Maitra, S., 2011. Who are the Indian Middle Class? A Mixture Model of Class Membership Based on Durables Ownership. Working Paper. https://www.sudeshnamaitra.com/Maitra-middle-class-mixture-May2011.pdf . Accessed 5 January 2022.
- McLachlan, G. J., Krishnan, T., 1996. The EM Algorithm and Extensions. Wiley Series in Probability and Statistics.
- McLachlan, G. J., Peel, D., 2000. Finite Mixture Models. Wiley Series in Probability and Statistics.

References

[1] Ablett, J., Baijal, A. et al., 2007. The Bird of Gold: The Rise of India's Consumer Market. McKinsey Global Institute Report, https://www.mckinsey.com/featured-insights/asia-pacific/the-bird-of-gold . Accessed 5 January 2022).

[2] Banerjee, A., Duflo, E., 2008. What is Middle Class about the Middle Classes Around the World? Journal of Economic Literature. 22, 3-28.

[3] Bar-Ilan, A., Blinder, A. S., 1988. The Life Cycle Permanent-Income Model of Consumer Durables. Annals of Economics and Statistics. 9, 71-91.

[4] Birdsall, N., 2010. The (Indispensable) Middle Class in Developing Countries; or, The Rich and the Rest, Not the Poor and the Rest. Center for Global Development Working Paper 207.

[5] Birdsall, N., Graham, C., Pettinato, S., 2000. Stuck In the Tunnel: Is Globalization Muddling the Middle Class? Brookings Institution Center on Social and Economic Dynamics Working Paper No. 14.

[6] Blischke, W. R., 1964. Estimating the Parameters of Mixtures of Binomial Distributions. Journal of the American Statistical Association. 59, 510-528.

[7] Britannica, The Editors of Encyclopaedia. "social class". Encyclopedia Britannica, https://www.britannica.com/topic/social-class . Accessed 5 January 2022.

[8] Deaton, A., Kozel, V., 2005. Data and Dogma: The Great Indian Poverty Debate. The World Bank Research Observer. 20, 177-199.

[9] Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B (Methodological). 39, 1-38.

[10] Easterly,W., 2001. The Middle Class Consensus and Economic Development. Journal of Economic Growth. 6, 317-335.

[11] Everitt, B. S., Hand, D. J., 1981. Finite Mixture Distributions Monographs on applied probability and statistics, London; New York: Chapman and Hall.

[12] Hastie, T., Tibshirani, R., Friedman, J., 2001. Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics.

[13] India Brand Equity Foundation (IBEF), 2005. Economic Indicators: India's Middle Class Dream Takes Shape. http://www.ibef.org/artdisplay.aspx?cat_id=391&art_id=5788 . Accessed 5 January 2022.

[14] Indiastat. http://www.indiastat.com .

[15] Maitra S., 2021. On the Theory and Measurement of Relative Poverty using Durable Ownership Data. Working Paper. https://www.sudeshnamaitra.com/Maitra_relative_poverty_Sep2021.pdf . Accessed 5 January 2022.

[16] Maitra, S., 2017. Class Matters: Tracking Urban Inequality in Post-Liberalization India using a Durables-based Mixture Model. Applied Economics Letters. 24(17), 1203-1207.

[17] Maitra, S., 2016. The Poor Get Poorer: Tracking Relative Poverty in India using a Durables-based Mixture Model. Journal of Development Economics. 119, 110-120.

[18] Maitra, S., 2011. Who are the Indian Middle Class? A Mixture Model of Class Membership Based on Durables Ownership. Working Paper. https://www.sudeshnamaitra.com/Maitra-middle-class-mixture-May2011.pdf . Accessed 5 January 2022.

[19] McLachlan, G. J., Krishnan, T., 1996. The EM Algorithm and Extensions. Wiley Series in Probability and Statistics.

[20] McLachlan, G. J., Peel, D., 2000. Finite Mixture Models. Wiley Series in Probability and Statistics.

[21] Pickering, G., 1968. High Blood Pressure. London: Churchill Ltd.

[22] Platt, R., 1963. Heredity in hypertension. Lancet. April, 899-904.

[23] Ravallion, M., 2010. The Developing World's Bulging (but Vulnerable) Middle Class. World Development. 38, 445-454.

[24] Senauer, B., Goetz, L., 2003. The Growing Middle Class in Developing Countries and the Market for High-Value Food Products. Department of Applied Economics Working Paper, University of Minnesota.

[25] Sridharan, E., 2004. The Growth and Sectoral Composition of India's Middle Class: Its Impact on the Politics of Economic Liberalization. India Review. 3, 405-428.