Who are the Indian Middle Class? A Mixture Model of Class Membership Based on Durables Ownership

Sudeshna Maitra

Department of Economics,

York University

4700 Keele Street

Toronto, ON M3J 1P3

Email: maitra@yorku.ca

Phone: 416-736-2100 Ext. 77052

Fax: 416-736-5987

Current Version: May 2011

Who are the Indian Middle Class? A Mixture Model of Class Membership Based on Durables Ownership

Sudeshna Maitra

Current Version: May 2011

ABSTRACT

A durables-based mixture model of class membership is used to identify the lower, middle and upper classes in urban India. Different from existing studies, this research does not identify classes by specifying exante income/expenditure cutoffs. Instead, classes are identified based on systematic differences in durables ownership. In our approach, classes are latent objects and the probability that any household belongs to one of the latent objects (classes) is estimable. These probabilities are used to further characterize classes in terms of class-specific socioeconomic characteristics. Using data from the National Sample Survey (1999-00), the richer probabilistic (latent) class definition that arises in this paper suggests that the urban middle class may be larger (and "poorer") than previously suspected. Since durable ownership is a marker for forward-looking households with sustainable living standards, the identified middle class arguably meets expectations of being drivers of economic growth, and should constitute the focus for policy.

Keywords: middle class; durable ownership; consumption; mixture model; EM algorithm

JEL classifications: O10, O12, O15, C14

1. Introduction

The 'middle class' has long been a topic of interest to economists, for a variety of reasons (Banerjee and Duflo, 2008). Birdsall, Graham and Pettinato (2000) describe the middle class as the "backbone of the market economy and democracy in most advanced societies." Easterly (2001) argues that countries which have a larger middle class tend to have higher growth rates. One way in which the middle class might boost growth and stability is through their 'middle class values' – such as emphasis on human capital accumulation and savings – which serve as valuable inputs to entrepreneurial activities. Another argument for why the middle class is crucial for growth emphasizes that it is from this class that new entrepreneurs often emerge; entrepreneurs who are characterized by a tolerance for delayed gratification and who engage in economic activities that generate employment and productivity growth in the rest of the economy. Yet another channel by which a larger middle class could spell higher growth is through the 'middle class consumer' who demands quality consumer goods and is willing to pay a higher price for better quality. This demand could potentially provide a 'big push' to investment in production and marketing and, in turn, provide an impetus for rising income levels (Murphy, Schleifer and Vishny (1989), Banerjee and Duflo (2008)).

The middle class has also been considered to be a group that is most likely to be susceptible to economic volatility. Ravallion (2010) argues for the vulnerability of the middle class in developing countries to the global economic crash of 2008, while Birdsall (2010) argues that the focus of inclusive growth in developing countries should move from the poor (and the rest) to the group that is "neither rich nor poor", viz. the middle class.

The size and characteristics of the *Indian* middle class have received considerable attention in recent years, for the above as well as other reasons. India's growth achievements since the 1990s have put the living standards of Indians under global scrutiny. While the economic literature has primarily focussed on poverty and inequality in India (see Deaton and Kozel (2005) for a review), the fortunes of the 'new Indian middle class' have received substantial attention in the media and in business journals, as their earning potential and spending habits have important implications for the global economy. Moreover, since India possesses a sixth of the world's population, its middle class arguably constitutes a significant portion of the global workforce as well as a substantial market for final products.

Who *are* the middle class, of whom we expect so much and about whom so much has been written? There is no clear consensus on this question in the small but growing literature that attempts to define and characterize the middle class. Existing studies (Banerjee and Duflo, 2008; Birdsall et al, 2000; Easterly, 2001; Ravallion, 2010; Birdsall, 2010; Ablett et al, 2007; NCAER, 2005; IBEF, 2005; Sridharan, 2004; Milanovic and Yitzhaki, 2002) have typically imposed income (or expenditure) cutoffs for the different classes, and then proceeded to examine the characteristics of the groups thus formed. This approach necessitates the use of several implicit assumptions – about who the different classes are and what their income (or expenditure) levels must be. Not surprisingly, research results are extremely sensitive to the definition of class boundaries. Also, the general lack of a consensus about which expenditure or income cutoffs to use suggests that it might be more appropriate to think of class 'boundaries' in a probabilistic sense, and that the income ranges corresponding to different classes should be thought of as being overlapping instead of being mutually exclusive (as suggested by the cutoffs approach)[1].

In this paper, I propose the use of a non-parametric method for identifying (latent) middle class households in India, that seeks to redress concerns raised

[1] 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 on developing nations.

about previous studies. The approach introduced here may be viewed as an approach that is 'dual' to that used in prior studies. In the existing literature, per capita expenditure (PCE) or income is used to define the classes, and various class characteristics are estimated conditional on the classes that are identified. Here, I propose to use a class characteristic – other than PCE – to define the classes, and then estimate PCE conditional on the estimated latent classes. In particular, I propose using *durable goods ownership* of households to help in identifying the classes.

Consumer durables assure a steady stream of utility in future periods, making their ownership a good measure of sustainable consumption 'standards' as well as a marker for the household's 'value for the future'. Both these characteristics – consumption standards and value for the future – feature prominently in broader discussions of what constitutes the 'middle class' and why they could be critical for economic growth. Therefore, to the extent that identifying such a 'middle class' is important for economic growth and for policy, durables ownership is recommended as an appropriate variable for class identification.

No doubt for the above reasons, the analysis of durable goods ownership of middle class households is a common feature of most studies of the middle class in India (and other developing countries). The durables that different studies focus on can be broadly placed in one of three categories: recreational goods (radios, TVs, audio/ video systems), electrical household appliances (refrigerators, washing machines, airconditioners) and transport goods (bicycles, two-wheelers, automobiles). This suggests a common notion running through existing research on what characterizes middle class households in developing countries – the ownership of recreational, household and transport durables – even when the researchers' explicit assumptions about the class-defining expenditure cutoffs are quite different. This common notion in the literature – coupled with the factors discussed in the previous paragraph – motivates my use of *durables ownership* (in the above three categories) to define the Indian middle class.

One other significant advantage of using durables ownership for identification is that it is less likely to be misreported (misremembered or 'rounded') than expenditure levels in the recall period.

I use a mixture model (McLachlan and Peel, 2000; Everitt and Hand, 1981) to model the distribution of durables ownership in urban India. I postulate the existence of three classes – lower, middle and upper – in a Three-Component Mixture Model framework, and focus on the total number of recreational, household and transport durables that a household owns at the time of interview. I then estimate the population shares and durables-ownership density functions of the three component classes such that the individual ownership densities can, in combination, explain the overall density of durables ownership.

One appeal of a mixture model is that it is non-parametric. There are no exante (external) assumptions about who populates the classes, apart from the fact that the classes are *different* in their patterns (densities) of durables ownership. The classes are assigned based on clusters in the durables ownership data, hence regularities in the data decide who are in what class rather than the researcher.

The empirical analysis involves maximum likelihood estimation, which can provide challenges in terms of parameter estimation and hypothesis testing for mixture models. 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 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) using iterated steps. The algorithm and its application to this analysis is described in Section 2.

The mixture model estimated by the EM algorithm yields a class structure – class-shares in the population and class-specific durables ownership densities – which can be used to characterize the upper, middle and lower classes. The solution is unique and provides an arguably more robust identification of the classes than has been obtained thus far. Also obtained are *probabilities* for each household that it belongs to each possible class. This contrasts with previous studies where households are assigned to classes with certainty. Finally, the PCE ranges corresponding to the different classes (Section 3.2) are obtained and shown to be overlapping.

The data come from the 55th Round of the Indian National Sample Survey (1999-00). I focus on the total of 12 durable items (recreational, household and transport goods) that a household may own at the time of interview. The middle class is largely perceived to be an urban phenomenon, hence I focus on the urban sub-sample of the National Sample Survey (NSS). However, the analysis may easily be extended to the rural sub-sample as well.

I find that lower, middle and upper class households constitute 20%, 62% and 18% of urban households, respectively. This implies an urban middle class of approximately 17% of all households in the population, given that 28% of all Indian households are urban (2001 census, Indiastat, http://www.indiastat.com). The mean number of goods owned by households in these classes are, respectively, 0.3, 3 and 6.3. Standard errors of estimates are small, supporting the existence of three classes with distinct ownership patterns of durables.

How do the mixture model estimates compare with those in existing studies of the Indian middle class? Notably, the per capita daily expenditure cutoffs used by Banerjee and Duflo (2008) and Ravallion (2010) to define the middle class are quite close to the expenditures of the middle class identified in this paper (by durables ownership). At the same time, the mixture estimates obtained here suggest larger middle and upper classes than are found by Sridharan (2004), Ablett et al (2007) and the NCAER (2005) and IBEF (2005) studies. Sridharan's (2004) estimate of the middle class is between 13% and 47% of urban households in 1998-99, depending on the breadth of his definition of middle class. Although these figures are considerably less than the mixture estimate of 62% (of urban households), the numbers are difficult to compare for two reasons. First, Sridharan has followed the NCAER approach and defined the classes by setting income cutoffs. Second, each of his definitions of middle class includes the 'High' income category[2] and excludes the 'Lower-Middle' income category. Including the 'Lower-Middle' group and excluding the 'High' group in the definition of middle class, yields an urban-share estimate of 68.5% (using Sridharan's estimates), which is much closer to 62%. This exercise demonstrates the ambiguity that has traditionally dominated the identification of the middle class, and essentially recommends the new method

[2] This is the highest income category in the analysis (Sridharan, 2004).

presented here for its intuitive approach to the issue.

Das (2001) makes a reference to the urban middle class as constituting 20% of the Indian population. While it is not clear how this figure has been arrived at, it is nevertheless close to my estimate of 17% (of total households).

The rest of the paper is organized as follows. The model is described in detail in Section 2. Section 3 presents results and Section 4 concludes the paper.

2. Methodology

2.1. Data and Definitions

The data used in this analysis comes from the urban sub-sample of the 55th Round of the Indian NSS (1999-00). The $48,924$ households in the sample are asked a battery of questions about their consumption habits and expenditures. For a list of durable items, they are asked to report how many pieces of each good are *in use* at the time of the interview. For each durable, I define 'ownership' as an indicator that at least one piece of the durable is *in use* in the household at the time of interview. The variable of interest Y is the *total* number of durable goods that a household 'owns' (by the above definition) at the time of interview. A mixture model hypothesizes that the density of Y is a weighted sum of densities of individual groups in the population. The goal is, therefore, to identify the distinct groups in the population such that their individual ownership densities or consumption patterns can, when weighted by estimated class-membership probabilities, explain the overall density of Y observed in the sample.

In the following analysis, I use the set of goods on which ownership information is available and that closely match those used in the literature[3]. These constitute 12 durable goods, which can be placed in three broad categories: recreational goods (record player/gramophone, radio, television, VCR/VCP, tape/CD player), electrical household appliances (electric fan, air conditioner, washing machine, refrigerator) and transport goods (bicycle, motor bike/ scooter, motor car/ jeep)[4].

Note from the definition of Y above that the intensity of durable ownership – how many pieces of a certain durable are in use – is *not* incorporated in how ownership is defined. Affluence is measured by the *variety* of services from durables owned, not the intensity of use of individual items. This is due to the fact that in-

[3]For instance, Banerjee and Duflo (2008) look at the ownership of radios, televisions and bicycles by the middle class. Senauer and Goetz (2003) consider these as well as refrigerators, washing machines and automobiles. The NCAER report (2005) on Indian markets consider, additionally, items such as electric fans and air conditioners. Studies have also explored the ownership of consumer electronic items (such as audio and video systems) by the middle class.

[4]The data does not allow us to discern the quality of durable goods in use in a household (e.g. models of cars or TVs or makes of audio/ video goods). But, to the extent that goods of higher quality (e.g. plasma TVs versus black-and-white TVs) are owned by households with *more* goods, ignoring durable-quality in the definition of Y is unlikely to impede an appropriate identification of the classes. Footnote 5 makes a similar point.

tensity of ownership may be higher in larger households not necessarily belonging to a higher class (larger households with more electric fans, for instance); hence including intensity of use in the definition of ownership may inappropriately ascribe higher affluence to larger households. Moreover, ignoring the intensity of use does not imply – for example – that households with four cars are treated identically to households with one car. What is important for identifying affluence is the *total* number of *distinct* durables; hence to the extent that households with four cars are also more likely to own a higher *total* number of distinct durables than households with one car, they are more likely to be identified (correctly) as more affluent.

Figure 1 presents the distribution of Y – the total number of the 12 durable goods that households own – in the sample[5]. Table 1 presents summary statistics for the ownership variables.

The bimodality and positive skewness of the distribution of Y in Figure 1 suggest that a mixture model may be an appropriate description of the latent class structure. The objective of this analysis is to identify the n distinct classes in the population such that their individual ownership densities or consumption

[5] Note that Y – the total number of durable items owned – incorporates a 'natural' weighting of different goods based on the associated level of affluence. For instance, cars occur in households with higher values of Y than radios, since on average cars occur in (more affluent) households with more total durables than do radios.

patterns can, in combination, explain a distribution like that in Figure 1.

Note that the application of a finite mixture model requires an assumption about n, the number of classes in the population. I argue that the appropriate number of classes is the *minimum* number of classes that can produce a good fit to the observed density of Y. Else, in the extreme case of allowing each household to be in a class of its own, a perfect fit could easily be obtained. In the present case, I show that a better fit is obtained when three classes are assumed than with two classes (see Section 3.1 and the appendix). Hence, a Three-Component Mixture Model is used to identify the three classes; henceforth referred to as the 'lower', 'middle' and 'upper' class, respectively. Details of the model and the estimation algorithm are provided in Sections 2.2 and 2.3.

Before proceeding to the formal model and estimation algorithm, however, it is useful to discuss why 'durables ownership' is used to identify the classes instead of per capita expenditure (PCE). I do this in the next subsection.

2.1.1. Why 'durables ownership'?

The notion of 'class' – particularly, the 'middle class' – embodies a complex combination of economic, social and political factors (Kapur, 2010; Singh, 2005). It is important, therefore, to recognize that the task of identifying 'classes' is not the same as that of measuring inequality (in income or expenditure) per se, although both share a common theme. Also, while PCE is a good general indicator of affluence, it is not clear why total expenditure should be associated with the (economic-growth-enhancing) characteristics that are expected of the 'middle class'. Hence, in order to answer the question – "who are the Indian middle class?" – we must use an identifying variable that can capture the *notion* of the 'middle class' as expressed in the broader literature.

I argue below that 'durables ownership' is able to capture two key features that resonate in broader discussions of what constitutes the 'middle class': (1) the importance of sustainable living (or consumption) 'standards' for defining a 'class' (Townsend, 1979); and (2) the lack of myopia (or synonymously, an increasing value for the future) embodied in the actions and decisions of the middle class (Banerjee and Duflo, 2008). The latter condition is clearly critical for driving middle class activities – such as entrepreneurship, human capital accumulation etc – that are deemed to enhance economic growth[6].

[6]The middle class is interesting because it is made up households who are neither rich nor poor. Therefore, a useful (if condensed) way of thinking about the middle class in the current context is as a group of households that has escaped from poverty and the associated preoccupation with myopic day-to-day survival. It is essential that this class of households be confident of not relapsing into poverty and confident of the positive impacts of their beliefs (economic, social or political) and actions on the personal growth that they aspire for (such as rising up the class order); in turn, driving economic growth. Consumption 'standards' (as opposed to consumption expenditure) encapsulates the confidence of class households of being able to ex-

Consumer durables are a store of utility that represent the stock component of household wealth rather than the flow component embodied in PCE. Moreover, the ownership of durables assures the realization of a stream of consumption utility in future periods. Both of these characteristics make durables ownership a good measure of consumption 'standards', since it represents a permanent, sustainable aspect of consumption (Bar-Ilan and Blinder, 1988). In contrast, measures based on total expenditure say relatively little about the 'standard' of living or its sustainability since they may include transitory components as well[7].

One may argue that a durable good may be acquired using transitory income, which would then make its ownership an inappropriate indicator of (permanent) living 'standards'. However, even if the above is true, it is reasonable to expect that a larger *total* number of durables *in use* – the measure of ownership used herein – is likely to represent a household with higher permanent income, and hence a higher sustainable standard of living. This recommends the use of the total

perience a sustained consumption pattern. Low myopia (or high value for the future) represents the confidence that current actions can and will lead to a better future.

[7] There is an implicit assumption that 'consumption' (and not political ideology, for instance) plays the key role in class definition (Kapur, 2010). This assumption has been made for two reasons. First, the consumption-oriented approach has been adopted in most of the economic literature on the middle class, where PCE is the primary indicator of class membership. Second, since the (growing) Indian middle class has been largely perceived as an outcome of higher rates of economic growth, it seems reasonable to believe that improving consumption standards may have constituted a direct force behind this phenomenon. Indeed, the increased materialism of the new Indian middle class and their aspirations to emulate modern western living standards – have been repeatedly mentioned in discussions of the issue (Varma, 1998; Singh, 2005).

number of durables in use as an indicator of higher permanent living standards, and hence of an increased probability of membership in a higher class.

Marking the distinction between classes identified by a 'permanent' consumption standard is important for policy since most of the benefits of a large middle class discussed in the literature arise from the sustainability of the middle class phenomenon, or the confidence of class members of being able to afford a permanently higher consumption standard than the lower class. Identifying who the Indian middle class are (and in future work, whether they are growing over time) by the 'permanent' criterion used herein is therefore a more pertinent question for policy, than that of identifying the middle class defined only in terms of overall consumption expenditure.

Consumption standards alone do not define a 'middle class'[8]. Our interest in the middle class is driven by their (expected) potential to engage in activities that fuel economic growth. Such activities require an important characteristic in the decision-maker: a lack of myopia, or an intrinsic value for the future. A key feature of durables is that ownership yields a stream of consumption utility

[8]This point, while being fairly obvious, needs to be underlined lest the current approach be construed as proposing a policy of distributing consumer durables to the lower class! Such a 'policy' would certainly achieve an expansion of a 'middle class' with a certain permanent consumption standard but would not necessarily enhance the size of the middle class who are likely to be engines of economic growth.

in future periods. Therefore, ownership of durables is arguably a marker for forward-looking households who hold some value for the future (Lastrapes and Potts, 2006); it is these households that are key for economic growth and hence for policy[9].

The ownership of consumer durable items has featured prominently in most studies of the middle class, no doubt for the above reasons. While the middle class is almost always defined using PCE-cutoffs, the patterns of durable ownership (specifically recreational durables, electrical household appliances and transport durables) among middle class households are invariably analyzed and documented, to throw light on who constitute the middle class. This indicates an implicit *definitional* linkage between durables-ownership and middle class status, even though the linkage has not been explicitly used to identify the class.

The approach adopted in this paper may, therefore, be considered a 'dual' approach whereby durables ownership is used to identify the classes and the PCE-ranges of the classes thus identified are subsequently explored. Indeed, the results obtained by this 'dual' approach (see Section 3.2) show that the PCE-range of the middle class identified by the durables-approach is comparable to the PCE-

[9] As before, note that while myopic households may occasionally be observed to own individual durable items, a higher *total* number of durables may reasonably be expected to represent a more forward-looking houshold with a higher value for the future.

cutoffs assumed by Banerjee and Duflo (2008) and Ravallion (2010). This suggests intuitively that the dual approach – of using durables ownership instead of PCE to identify the classes – is able to identify a middle class that corresponds to existing researchers' notions about the same. In addition, the mixture approach allows the data to determine the *distribution* of middle class households over the relevant PCE-range, instead of assuming that *every* household in this PCE-range belongs to the middle class with certainty, as in the cutoffs-based approach (Section 3.2 provides a further discussion of this point).

Furthermore, the fact that 'durables ownership' – not PCE – is used to identify the households in the first place suggests that the estimated distribution of middle-class households (over the relevant PCE-range) is driven by a direct measure of middle class characteristics: sustainable consumption standards and an implicit value for the future. Hence, the households identified as 'middle class' in the current approach are arguably closer to *being* 'middle class' – than those identified by PCE-based criteria – by the notion that appears in broader discussions of 'class'.

A final, technical reason for using durables ownership for class identification is that durables ownership data are relatively free of reporting errors such as recall errors and rounding. The durables approach is especially amenable to comparisons of the Indian middle class over time using the National Sample Survey (NSS). The NSS recall periods for reporting expenditures were altered in the 1999-2000 round, leading to widespread concerns that expenditures reported in the later surveys may suffer from a systematic recall bias (see the poverty literature summarized in Deaton and Kozel, 2005). Durables ownership information – measured by whether or not certain durables are in use at the time of the survey – is not affected by a change in recall periods and using these would consequently enable reliable comparisons of class characteristics over time. Therefore, to the extent that there is an innate dynamic element in the answer to the question "who are the Indian middle class" and it is worth exploring this element in future work, there is a clear rationale for proposing an approach that uses durables-ownership information for identification.

I now turn to a formal presentation of the three-component mixture model used here.

2.2. The Three-Component Mixture Model

Consider 12 durable goods and let Y represent the total number of these goods that a household owns at the time of interview, $Y \in \{0, 1, 2 \ldots 12\}$. Households can belong to one of three classes – 1, 2 or 3 – which are defined by the pattern of durables ownership of members. 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 ($= 1, 2, 3$) 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[10].

Note that the class-specific probability of ownership p_i – which may be interpreted as the probability that a class-i household owns a representative durable good – is assumed to be the same for each good. This is for the following two reasons. First, allowing the probability to vary by class as well as good, viz. p_{ij} ($i = 1, 2, 3;\ j = 1, ..., 12$)[11] would make the mapping of parameters $\{(\pi_i, p_{ij}), i = 1, 2, 3; j = 1, 2, ..., 12\}$ to class $\{Lower,\ Middle,\ Upper\}$ less transparent since there is no longer a clear and intuitive ordering of p_i's that defines the classes. In other words, we would now have to choose some external criterion to compare the vector $(p_{i1}, ...p_{i12})$ across classes $i = 1, 2, 3$ and determine which

[10] Allowing dependence in the ownership of different goods would necessitate several additional assumptions on the nature of dependence. Derivation of the density functions ϕ_i in these cases becomes very complex.

[11] Here j represents a particular durable good, not the total number of durables. Since there are 12 durables in the analysis, j can take values $1, 2, ..., 12$.

of these is the lower, the middle and the upper class. Second, it is not the focus of the current analysis to explore the characteristics of the goods j per se (viz. necessary/ luxury items) but to identify the three classes represented by distinct patterns of affluence. A fundamental premise of the current approach is that affluence (and therefore class status) is measured by the *total* number of durables owned. Assuming p_i (and not p_{ij}) provides the simplest tractable framework within which to exploit this premise and generate a transparent mapping of parameters to class[12].

The probability of obtaining an observation y in the sample is given by:

$$P(y; \pi_1, \pi_2, p_1, p_2, p_3) = \pi_1 \phi_1(y; p_1) + \pi_2 \phi_2(y; p_2) + (1 - \pi_1 - \pi_2) \phi_3(y; p_3) \quad (1)$$

where π_i represents the probability that the household belongs to class i and $\phi_i(y; p_i)$ represents the (binomial) probability that a class-i household owns y durables. This is a Three-Component Mixture Model (McLachlan and Peel, 2000; Everitt and Hand, 1981).

[12] Note also that postulating a mixture model that allows $p_{ij}(i = 1, 2, 3; j = 1, 2, ..., 12)$ involves the estimation of 38 parameters $(\pi_1, \pi_2, \{p_{ij}\}_{i=1,2,3}^{j=1,...,12})$. It is hard to establish the identifiability of such a model.

2.2.1. Identifiability and Observational Equivalence

Before attempting to estimate the binomial mixture model in (1), it is necessary to establish that the model is identifiable. While binomial mixtures in the parameter p need not be identifiable in general (Teicher, 1961), a well-known paper by Blischke (1964) shows that a necessary and sufficient condition for identifiability is $n \geq (2r-1)$, where n is the binomial parameter denoting the number of trials and r is the number of components in the mixture. In the current application, $n = 12$ (the number of durables) and $r = 3$ (the number of classes), so the condition for identifiability is easily satisfied. Hence the model (1) is identifiable.

Note also the issue of observational equivalence known to characterize mixture models in general. This means – for example – that there is no difference observationally, between 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 equivalance 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.

2.2.2. Estimation

Having established identifiability, we now proceed to estimation of the mixture model. From (1), the likelihood function can be written as

$$L(y; \pi; p) = \prod_{j=1}^{N} [\pi_1 \phi_1(y_j; p_1) + \pi_2 \phi_2(y_j; p_2) + (1 - \pi_1 - \pi_2)\phi_3(y_j; p_3)]$$

where subscript j denotes the household, $j = 1, 2, ..., N$. The log likelihood function is then:

$$\log\ L(y; \pi; p) = \sum_{j=1}^{N} \log\ [\pi_1 \phi_1(y_j; p_1) + \pi_2 \phi_2(y_j; p_2) + (1 - \pi_1 - \pi_2)\phi_3(y_j; p_3)] \quad (2)$$

It is hard to obtain closed-form expressions for maximum likelihood estimates of the parameters in (2). The 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 2.3. 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.

2.3. Implementation of the Expectations Maximization (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.

$$\begin{aligned}
\delta_{1j} &= 1 \text{ if household } j \text{ belongs to class } 1 \\
&= 0, \text{ otherwise} \\
\delta_{2j} &= 1 \text{ if household } j \text{ belongs to class } 2 \\
&= 0, \text{ otherwise}
\end{aligned}$$

If class memberships (δ_1, δ_2) were *not* latent variables, then the likelihood and log-likelihood functions could be written as

$$L_{EM}(y;\pi;p) = \prod_{j=1}^{N} \{\pi_1\phi_1(y_j;p_1)\}^{\delta_{1j}} \{\pi_2\phi_2(y_j;p_2)\}^{\delta_{2j}} \{(1-\pi_1-\pi_2)\phi_3(y_j;p_3)\}^{(1-\delta_{1j}-\delta_{2j})}$$

$$\begin{aligned}
\log\ L_{EM}(y;\pi;p) &= \sum_{j=1}^{N} [\delta_{1j} \log\ \{\pi_1\phi_1(y_j;p_1)\} + \delta_{2j} \log\{\pi_2\phi_2(y_j;p_2)\} \\
&\quad + (1-\delta_{1j}-\delta_{2j}) \log\{(1-\pi_1-\pi_2)\phi_3(y_j;p_3)\}]
\end{aligned} \tag{3}$$

It would be easy to find closed-form expressions for maximum likelihood parameter estimates from (3), if class memberships (δ_1, δ_2) were known. Since class memberships are unknown, 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 (3) 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 $(y_1, y_2, ..., y_N)$. Since class memberships are binary, $\gamma_i^{(k)}$ is also the estimated probability that a household belongs to class i, conditional on the data.

$$\begin{aligned} \gamma_{ij}^{(k)} &= E(\delta_{ij} | (y_1, y_2, ..., y_N; \pi_1^{(k-1)}, \pi_2^{(k-1)}, p_1^{(k-1)}, p_2^{(k-1)}, p_3^{(k-1)}) \qquad (4) \\ &= \frac{\pi_i^{(k-1)} \phi_i(y_j; p_i^{(k-1)})}{\pi_1^{(k-1)} \phi_1(y_j; p_1^{(k-1)}) + \pi_2^{(k-1)} \phi_2(y_j; p_2^{(k-1)}) + (1 - \pi_1^{(k-1)} - \pi_2^{(k-1)}) \phi_3(y_j; p_3^{(k-1)})} \end{aligned}$$

$i = 1, 2, 3.$

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 (3), when (δ_1, δ_2) are replaced by their expected values conditional on the data.

$$\pi_i^{(k)} = \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij}^{(k)} \tag{5}$$

$$p_i^{(k)} = \frac{1}{12}\left[\frac{\sum_{j=1}^{N}\gamma_j^{(k)}y_j}{\sum_{j=1}^{N}\gamma_j^{(k)}}\right]$$

$i = 1, 2, 3.$

4. Iterate steps 2 and 3 (the E and M steps) till convergence is obtained.

As output, the EM algorithm yields the following estimates:

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

The ownership probabilities $\widehat{p}_i$ and the corresponding class-specific densities $\phi_i(y;\widehat{p}_i)$ answer our motivating question – who are the Indian middle class? – by identifying the distinct ownership patterns of the different classes. Moreover, the estimates of the unconditional probabilities $\widehat{\pi}_i$ – interpretable as estimates of class shares – tells us the size of the urban middle class in India. Finally, the estimated (conditional) probabilities of class membership, $\widehat{\gamma}_{ij}$, along with $\widehat{\pi}_i$ and $\widehat{p}_i$, enable an 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.

The next section presents the results.

3. Results and Discussion

3.1. Estimates

The estimates produced by the EM algorithm are presented in Table 2 and Figures 2 to 4.

The numbers in column *(1)* of Table 2 represent the population share of each class, $\widehat{\pi}_i$. 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.

Column *(2)* reports estimates of the probability parameter $\widehat{p}_i$ 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[13].

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 2. The lower, middle and upper classes are found to own, on average, 0.3, 3 and 6.3 goods, respectively.

Figure 2 plots the binomial density functions ϕ_i at the estimated parameters $\widehat{p}_i$ ($i = 1, 2, 3$). Classes 1, 2 and 3 are the lower, upper and middle classes, re-

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

spectively. The density of the lower class peaks at 0 durables, whereas that of the middle and upper classes peak at 3 and 6 durable goods, respectively.

Figure 3 plots the actual relative frequency of observations (Y) in the data along with the predicted values. The figure demonstrates a very good fit to the data. As an analytical exercise, a Two-Component (two classes) Mixture Model was fitted to the data by EM. The results are presented in the appendix. The fit is clearly worse than that of the Three-Component Model. Hence, three appears to be the minimum number of classes that provides a good fit to the data.

Figure 4 plots the probabilities $\widehat{\gamma}_i$ that a household belongs to different classes i $(= 1, 2, 3)$ conditional on the number of durables owned. For example, 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 2). Hence, unlike in previous studies, the current approach places households in different classes with a probability rather than with certainty.

3.2. Class Characteristics: A Descriptive Analysis

The current approach assigns households to different classes with a probability rather than with certainty. However, using the estimates obtained herein, it is possible to estimate the number of observations of each value of Y that belongs to each class. Based on this computation, I randomly assign households to classes. Here is an example of how the assignment is performed. Suppose we estimate $\widehat{\gamma}_{10} = 0.6, \widehat{\gamma}_{20} = 0.1$ and $\widehat{\gamma}_{30} = 0.3$. This means that a household that owns none of the twelve durable goods (i.e. $Y = 0$) belongs to class 1 with probablity 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 $Y = 0$. I then randomly assign 60 of these 100 households with $Y = 0$ to class 1, 10 to class 2 and 30 to class 3. The same procedure is followed to randomly assign households to classes for each other value of Y.

Assigning a class to each households allows a descriptive analysis of the characteristics of each class. I focus on the durables ownership patterns for specific goods as well as a host of socioeconomic characteristics. The results – which further illuminate who are the Indian middle class – are presented in Tables 3-4 and Figures 5-12 and discussed below.

Tables 3(a) – (b) and Figures 5(a) – (b) demonstrate 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[14]. Of these, electric fans and televi-

[14] This could be partly attributable to the fact that, among the 12 goods considered, there are

sions are most popular among the top two classes, whereas fans and bicycles are most popular among the lower class.

Table 4 reports the per capita monthly expenditures of households in each assigned class. The first point to note is that the ranges of PCE corresponding to different classes are overlapping rather than mutually exclusive, as suggested by the 'cutoffs' approach used in the literature. Moreover, the PCE ranges obtained here suggest lower income cutoffs for the different classes than have been used in many prior studies on India. As an illustration, consider the following approximate calculation. At a household savings rate of 28% (Ablett et al, 2007) and using the mean class-specific household sizes in the sample (see Table 4), I find median annual household incomes to be Rs. 41354.16 ($ 3432 at 2005 PPP[15]), 58420 ($ 4849 at 2005 PPP) and 104465 ($ 8671 at 2005 PPP) for the lower, middle and upper classes respectively. The NCAER study (2005) places the 'middle class' in the annual-household-income range of Rs. $200,000 - 1,000,000$ in 2001-02. The class immediately below the middle class – viz. 'aspirers' – are also placed in an income range that appears too high, viz. $Rs.90,000 - 200,000$, annually[16].

more recreational and household goods (5 and 4, respectively) than there are transport goods (3).

[15] Using Indian CPI inflation of 1.215 between 2000 and 2005 (Indian Labour Bueau) and a PPP conversion rate of Rs. 14.67 per US$ (ICP 2005).

[16] The NCAER study (2005) divides households into 4 classes: Deprived, Aspirers, Middle Class and Rich.

Notably, however, the expenditure cutoffs used by Banerjee and Duflo (2008) and Ravallion (2010) are well able to capture the range of middle class expenditures found here. Banerjee and Duflo 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 middle class has daily per capita expenditures of $\$2-\$13$ at 2005 PPP. The $1st$ and $99th$ 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)[17]. Also, the mean number of durables owned 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 similarity in the ranges of expenditures obtained here and those assumed by Banerjee and Duflo (2008) and Ravallion (2010) is heartening since it suggests

[17] 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) definition of the middle class (those lying between the 20th and 80th 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) definition (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 definition of the "indispensable middle class") would exclude most of the middle class identified herein.

intuitively that not much is lost by using durables (instead of PCE), to identify who the middle class are. However, 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 the middle class, whereas the mixture aproach identifies a middle class whose expenditures are *distributed* over a PCE-range. Consequently, 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 middle-class-expenditures is comparable across approaches. As an example, consider the estimate of the *size* of the middle class as a proportion of urban households. Using Banerjee and Duflo's definition in the current 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 herein suggests a middle class that comprises 62% of urban households[18].

Figure 6 plots the education levels of the household head, by class. The lower

[18] It is not informative to compare the size-estimates implicit 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).

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. Despite the mean proportion of literate middle-class-household members being 77% (see Table 4), this finding is somewhat surprising given the perception of the middle class as white-collar workers. However, the phenomenon would be consistent with an environment of social mobility characterized by a large influx of lower class members into the middle class. Repeating the EM analysis for other rounds of the NSS could provide further insight into this phenomenon.

Figure 7 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 8 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%).

Figures 9 and 10 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.

Finally, Figures 11 and 12 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). Likewise, Scheduled Castes and Tribes form a larger component of the lower than the middle and upper classes.

4. Summary and Conclusion

I propose the use of a mixture model as a robust method for identifying and estimating the size of the urban middle class in India, with classes defined by their distinct patterns of durables ownership.

Durables ownership – which assures a steady stream of consumption utility in future periods – is able to capture two key features of the 'middle class' that resonate in broader discussions of 'class': (1) the sustainability of consumption 'standards' and (2) the household's value for the future. To the extent that a sustainable middle class comprising forward-looking households is essential for economic growth, the middle class identified herein is the relevant category of households that should constitute the focus of policy.

Using a Three-Component Mixture Model and data on the total number of durables owned by households (NSS, 55th Round, 1999-00), I obtain estimates of the urban-population shares of the three classes (lower, middle and upper) as well as the probability that a household belonging to each class will own a durable good. The estimates are precise, with small standard errors, supporting the existence of three distinct durables ownership patterns – hence, three distinct classes – in the Indian urban population in 1999-2000.

The range of per capita expenditures of the middle class identified herein are very close to the class-defining cutoffs used by Banerjee and Duflo (2008) and Ravallion (2010). But the magnitudes of the share estimates obtained here indicate a larger urban middle and upper class (62% and 18%, respectively) than were found in many previous studies on India (Ablett et al, 2007; NCAER, 2005; IBEF, 2005; Sridharan, 2004). However, these previous studies have relied on several exante assumptions about who constitutes the classes, to which their results appear to be sensitive. The approach used here is free from such arbitrary assumptions and allows an identification of the classes based on their *distinct* durables ownership patterns. The solution obtained is unique.

The contribution of this paper is, therefore, twofold. First, by using durables ownership data for identification, we are able to identify a middle class that arguably conforms to the notion of the economic-growth-enhancing 'middle class' emphasized in the literature. It is this group of households who should constitute the focus for policy; yet it is not clear how well traditional expenditure-based approaches are able to identify them. Second, the mixture approach yields a richer probabilistic class definition than that obtained from cutoffs-based approaches. In the mixture approach, the classes are identified using regularities in the data rather than arbitrary researcher-driven assumptions. These factors recommend the use of a durables-based mixture approach to robustly identify the classes.

Acknowledgements: I am especially grateful to Barry Smith for his invaluable insights and encouragement. I thank The Conference Board, New York for support and Bart van Ark for any useful discussions. My thanks also to Alessandro Tarozzi for his suggestions. All remaining errors in the paper are my own.

References

[1] Ablett, J., Baijal, A. et al., 2007. The Bird of Gold: The Rise of India's Consumer Market. McKinsey Global Institute Report (available at: http://www.mckinsey.com/mgi/publications/india_consumer_market/index.asp; accessed Dec. 10, 2010).

[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] Das, G., 2001. India's Growing Middle Class. The Globalist.

[8] Deaton, A., Kozel, V. 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 Statisticial 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. (available at: http://www.ibef.org/artdisplay.aspx?cat_id=391&art_id=5788; accessed Dec. 10, 2010).

[14] Indiastat (available at: http://www.indiastat.com).

[15] Kapur, D., 2010. The middle class in India: a social formation or political actor? in Go, J. (Ed.) Political Power and Social Theory, Volume 21, Emerald Group Publishing Limited, pp.143-169.

[16] Lastrapes, W. D., Potts, T. B., 2006. Durable Goods and the Forward-Looking Theory of Consumption: Estimates Implied by the Dynamic Effects of Money. Journal of Economic Dynamics and Control. 30, 1409-1430.

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

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

[19] Milanovic, B., Yitzhaki, S., 2002. Decomposing World Income Distribution: Does the World Have a Middle Class? Review of Income and Wealth. 48, 155-178.

[20] Murphy, K. M., Schleifer, A., Vishny, R., 1989. Industrialization and the Big Push. Journal of Political Economy. 97, 1003-1026.

[21] NCAER Report, 2005. The Great Indian Market. (Preview slides available at: http://www.ncaer.org/downloads/PPT/TheGreatIndianMarket.pdf; accessed Dec. 10, 2010).

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

[23] 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.

[24] Singh, N., 2005.The Idea of South Asia and the Role of the Middle Class. Working Paper, University of California, Santa Cruz. (available at: http://mpra.ub.uni-muenchen.de/1277/1/MPRA_paper_1277.pdf; accessed May 30, 2011).

[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.

[26] Teicher, H., 1961. Identifiability of Mixtures, The Annals of Mathematical Statistics. 32, 244-248.

[27] Townsend, P., 1979. Poverty in the United Kingdom: a survey of household resources and standards of living. Berkeley: University of Berkeley Press.

[28] Varma, P. K., 1998. The Great Indian Middle Class, Viking Penguin India, New Delhi.

Table 1: Summary Statistics, Urban Sub-sample, NSS 1999-00, N = 48,924 households

Variable	Mean	Std. Dev.	Min.	Max.	Notes
Total number of goods 'owned' (Y)	3.06	2.33	0	12	Variable Used in Estimation
If household 'owns':					'1' if household owns at least one piece of the item
Record Player/ Gramophone	0.02	0.13	0	1	
Radio	0.36	0.48	0	1	
Television	0.60	0.49	0	1	Recreational Goods
VCR/ VCP	0.05	0.21	0	1	
Tape/ CD Player	0.30	0.46	0	1	
Electric Fan	0.67	0.47	0	1	
Air Conditioner	0.12	0.32	0	1	Household Goods
Washing Machine	0.10	0.30	0	1	
Refrigerator	0.25	0.43	0	1	
Bicycle	0.37	0.48	0	1	
Motor bike/ Scooter	0.20	0.40	0	1	Transport Goods
Motor car/ Jeep	0.03	0.17	0	1	
'Owns' at least one durable good	0.83	0.37	0	1	
'Owns' at least one recreational good	0.72	0.45	0	1	
'Owns' at least one household good	0.69	0.46	0	1	
'Owns' at least one transport good	0.50	0.50	0	1	
Total number of recreational goods 'owned'	1.32	1.08	0	5	
Total number of household goods 'owned'	1.13	1.08	0	4	
Total number of transport goods 'owned'	0.60	0.68	0	3	
Per Capita Monthly Household Expenditure	1018.73	1535.32	17	205987	48, 921 obs.

Table 2: Lower, Middle and Upper Classes in the Urban Sub-sample, Indian NSS, 55th Round (1999-00), N = 48,924 households

Category (Class)	**Mixture Estimates (Std. Error)**		
	(1) Share of Urban Population	*(2)* Probability of Owning a Good	*(3)* Mean No. of Goods (of 12)[a]
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)

[a] 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)

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

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)

Fig. 1: Distribution of Y

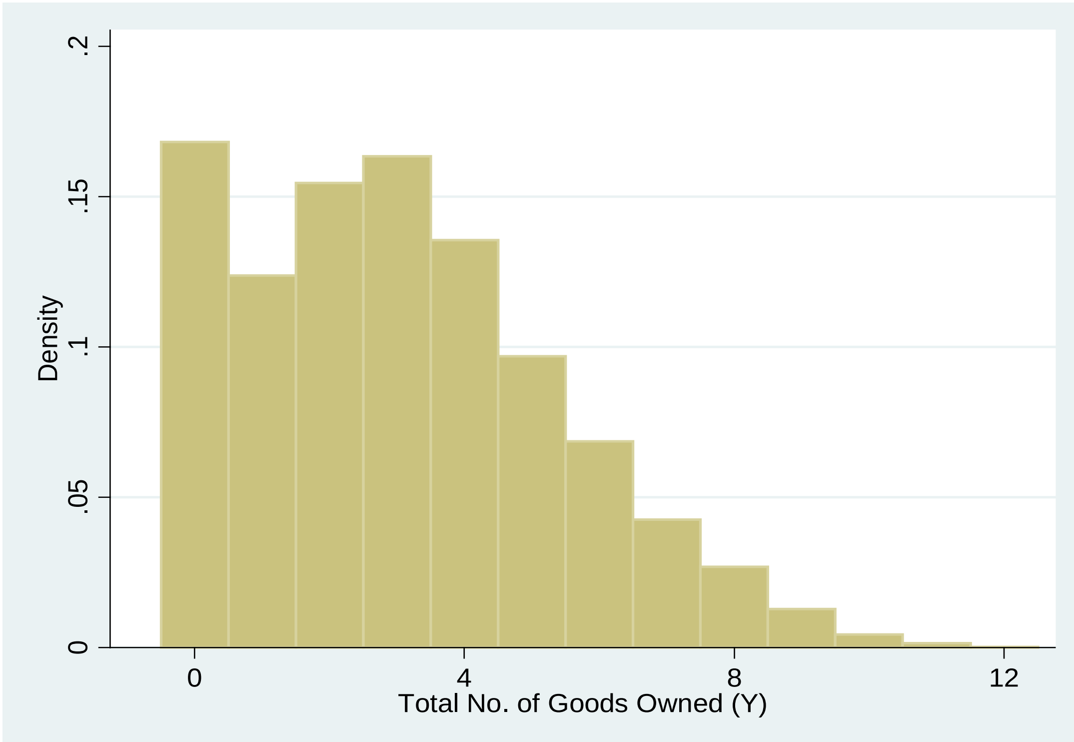


Fig. 2: EM-Estimated 'Density' Function of *Y*, by Class

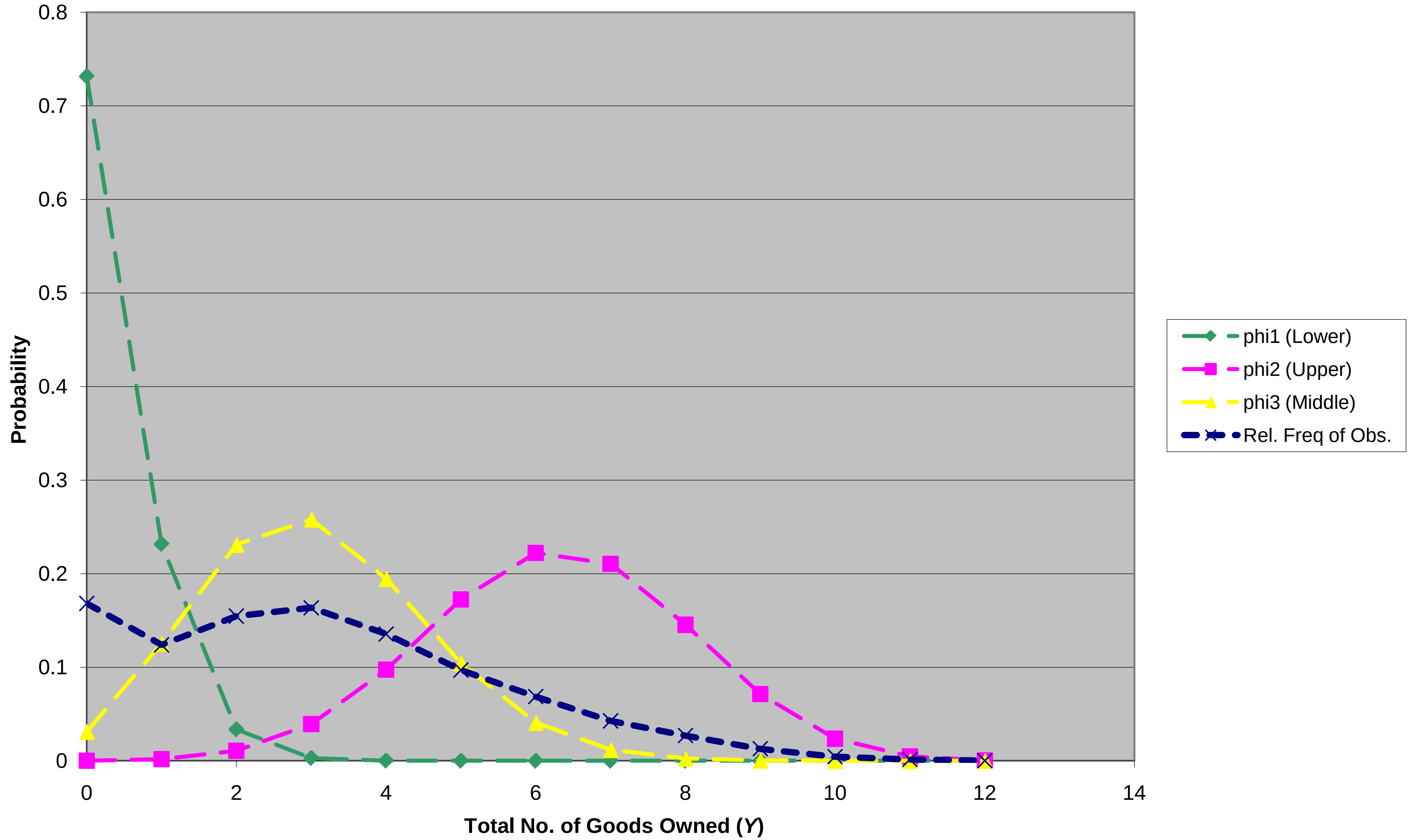


Fig. 3: Actual vs. EM-Predicted Distribution of *Y*

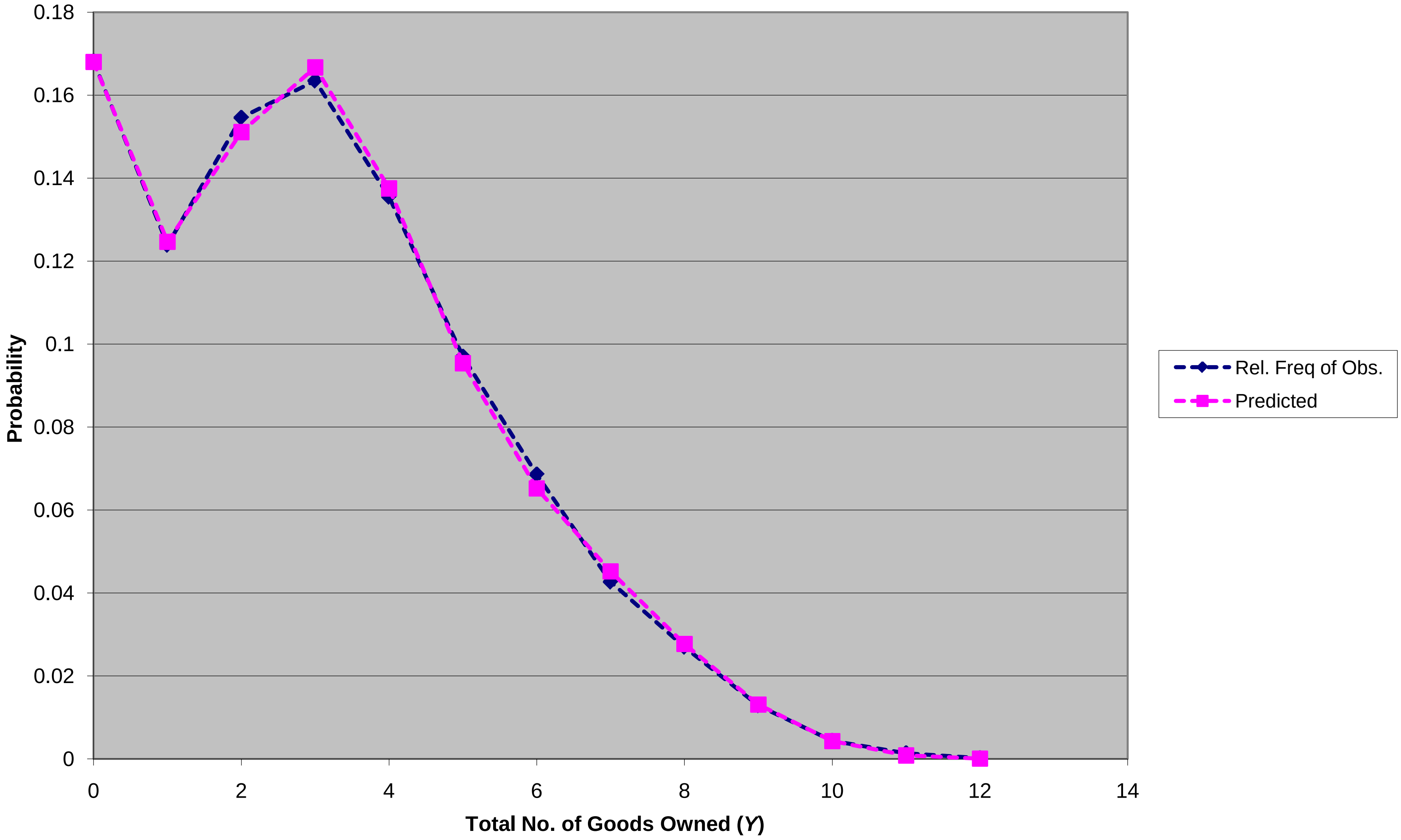


Fig. 4: EM-Estimated Probability of Belonging to Each Class

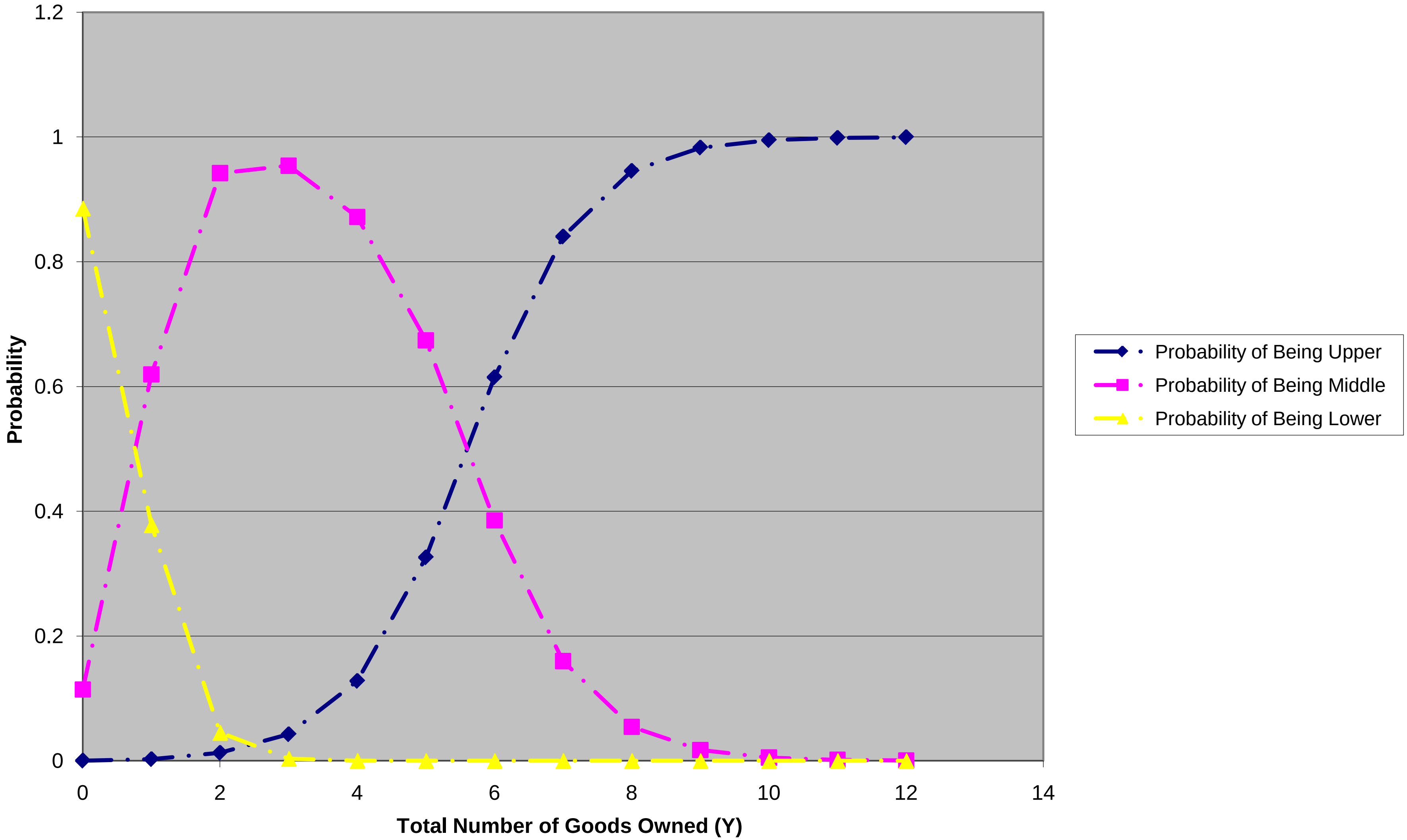


Fig. 5(a): Ownership by Durable Categories by Class, Urban Sub-sample, NSS 1999-00

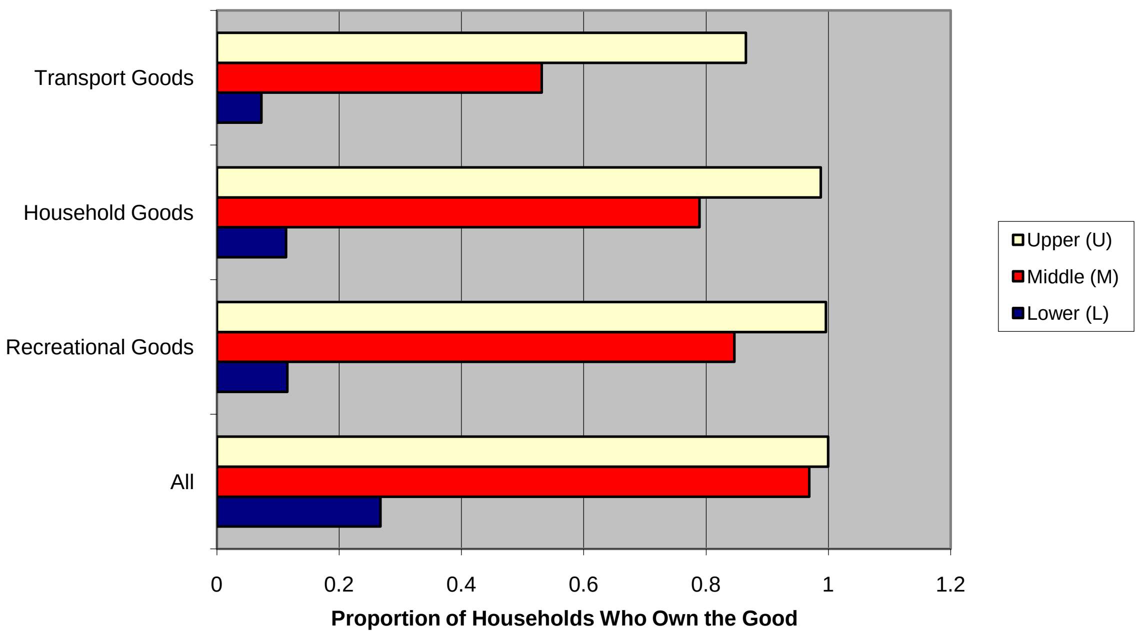


Fig. 5(b): Ownership of Individual Goods by Class, Urban Sub-sample, NSS 1999-00

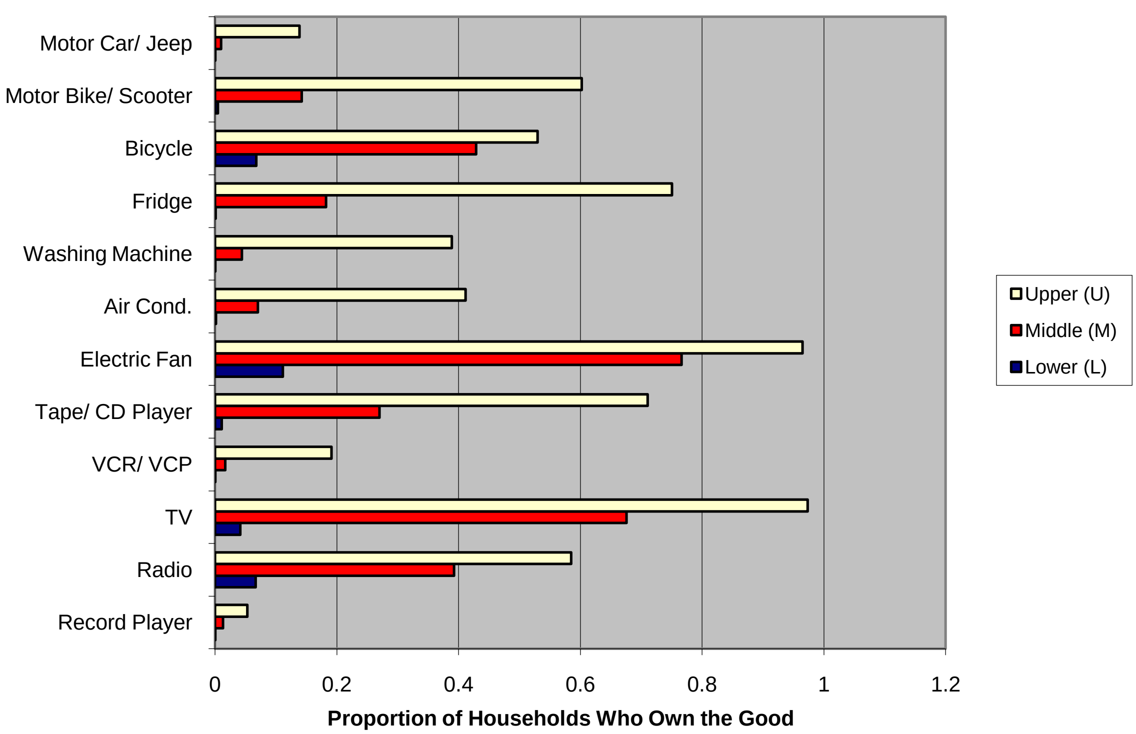


Fig. 6: Education of Household Head, by Class

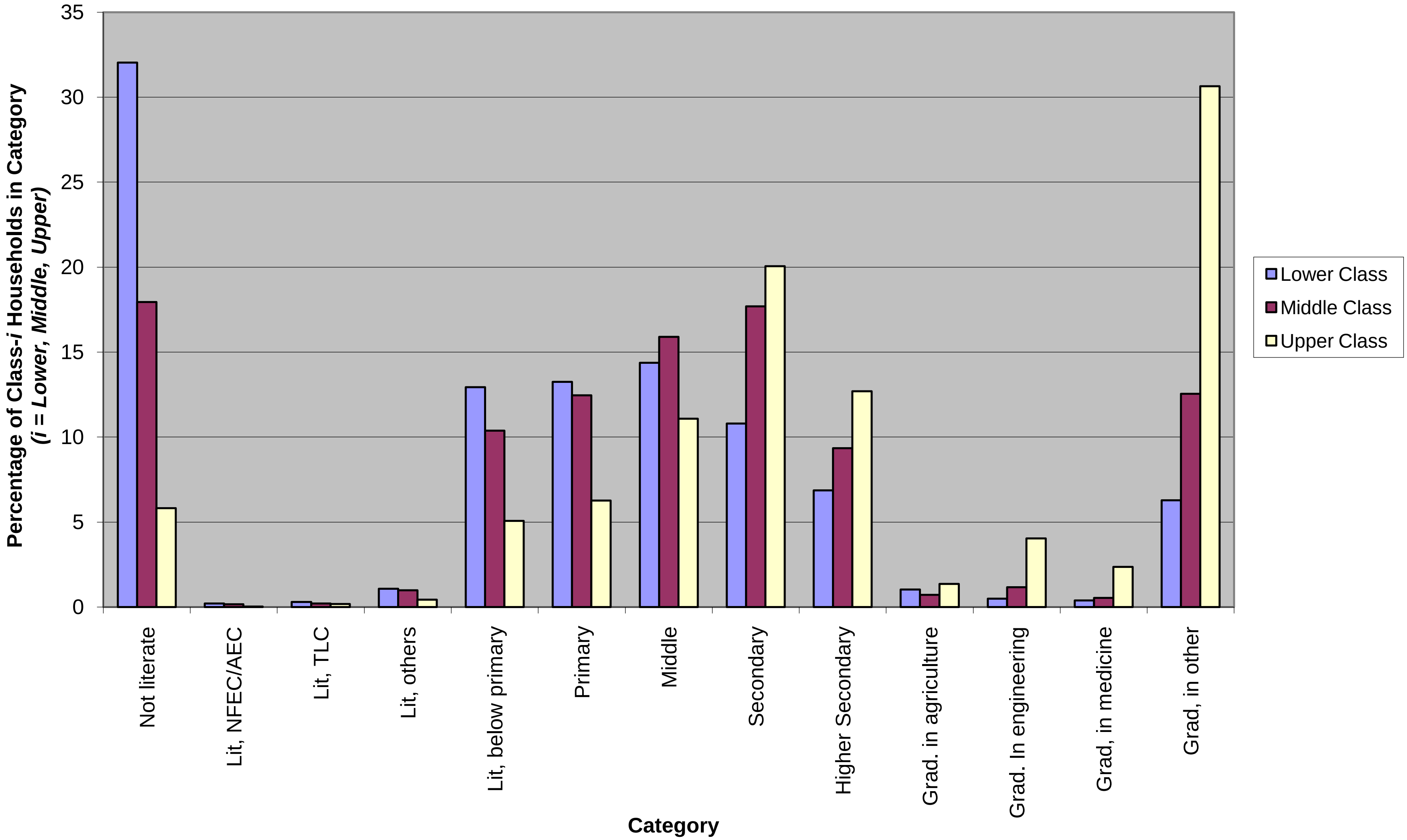


Fig. 7: Type of Employment, by Class

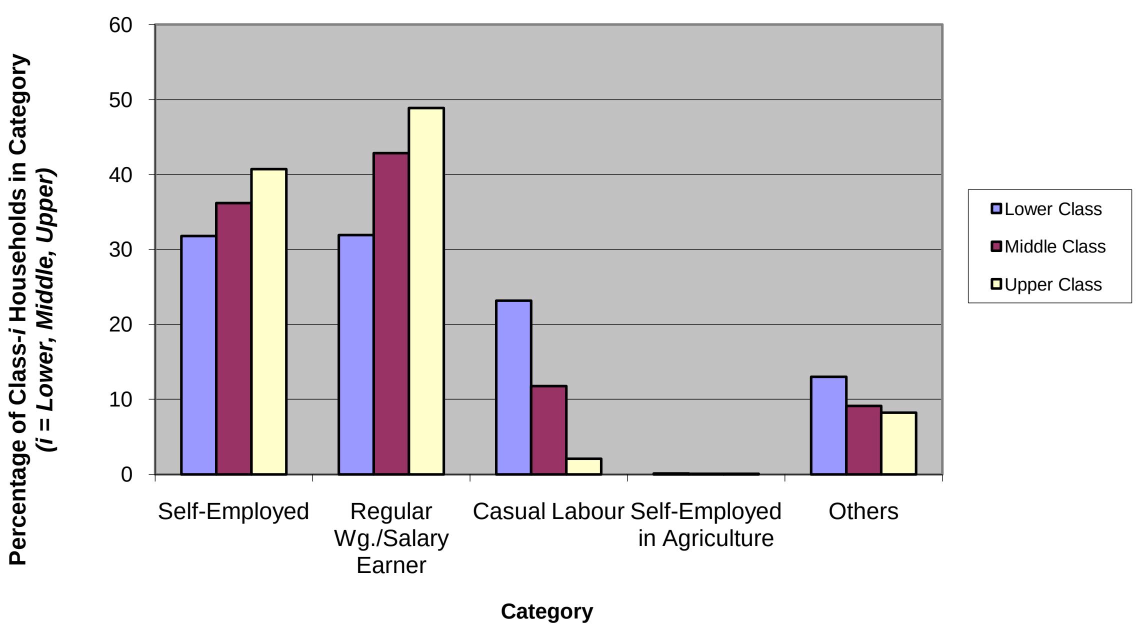


Fig. 8: Land Ownership & Source of Income, by Class

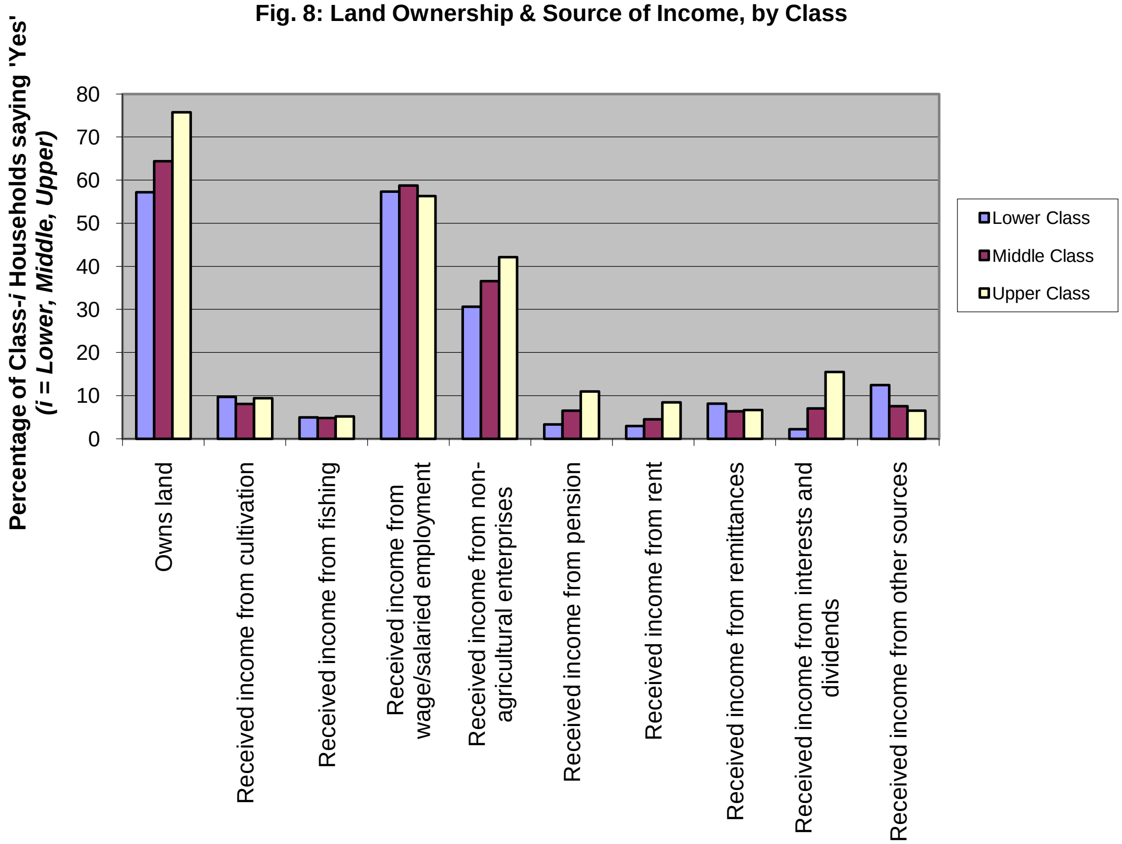


Fig. 9: Primary Source of Energy Used for Cooking, by Class

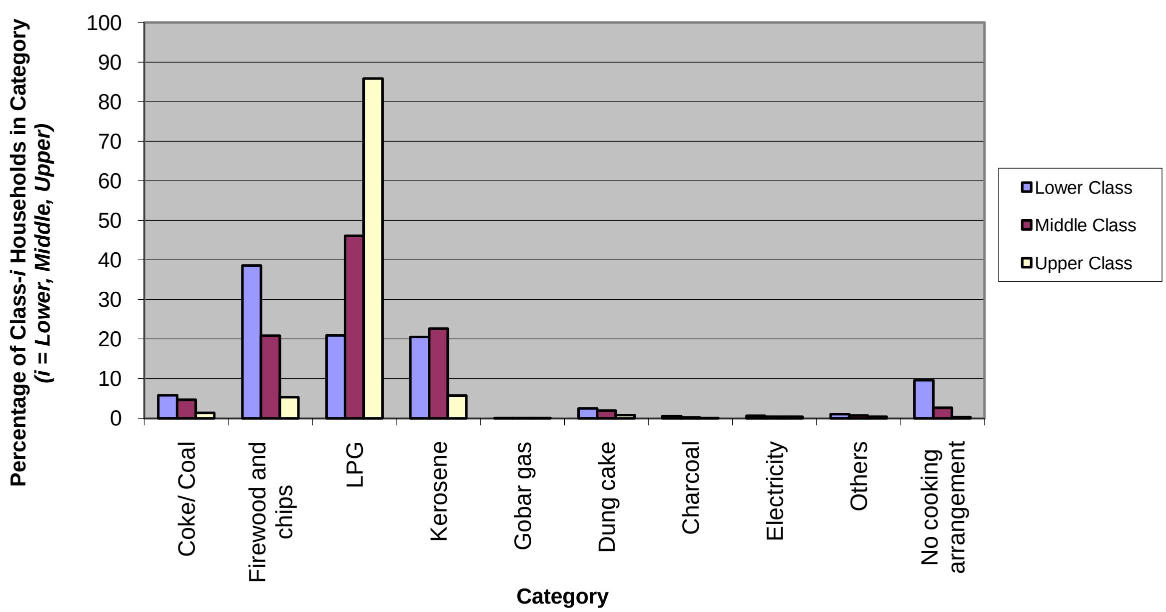


Fig. 10: Primary Source of Energy Used for Lighting, by Class

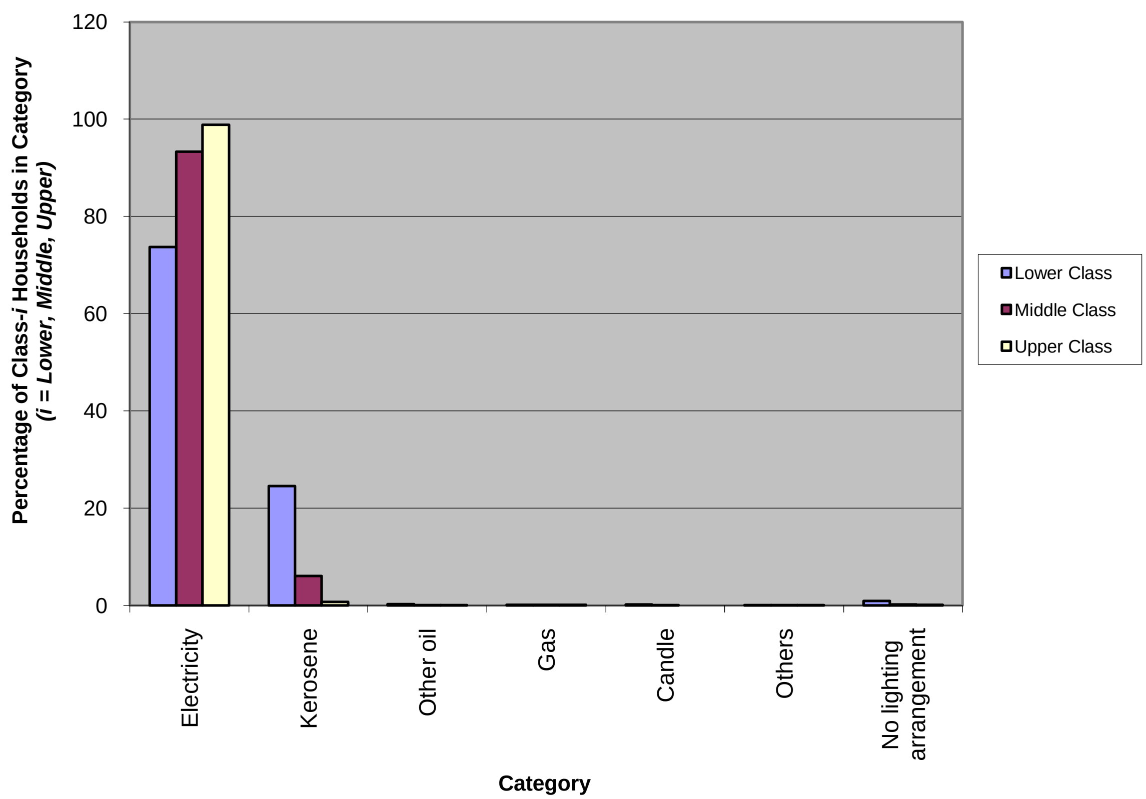


Fig. 11: Religion, by Class

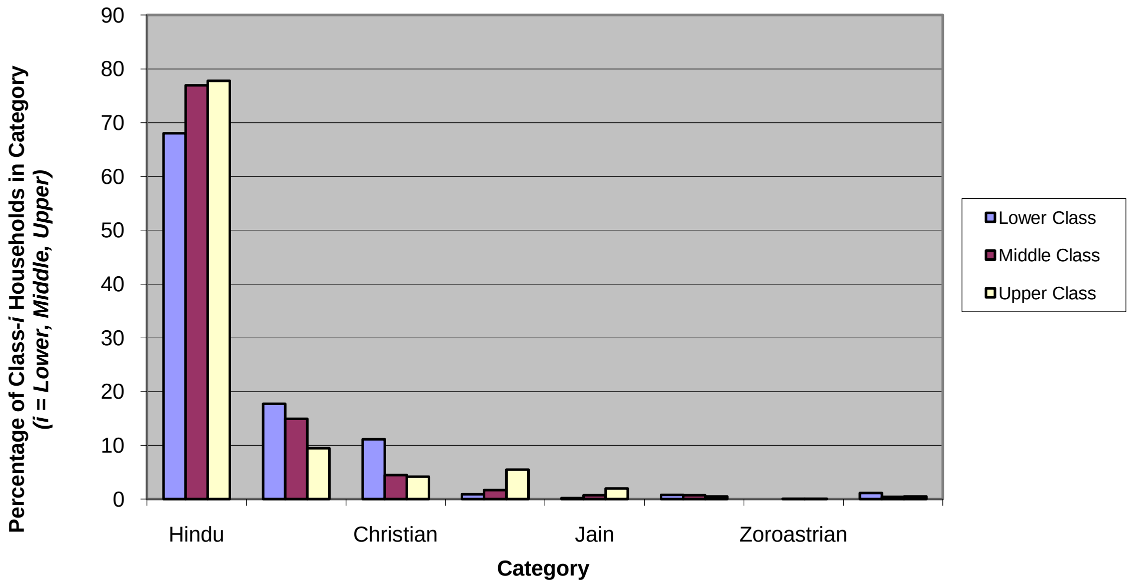
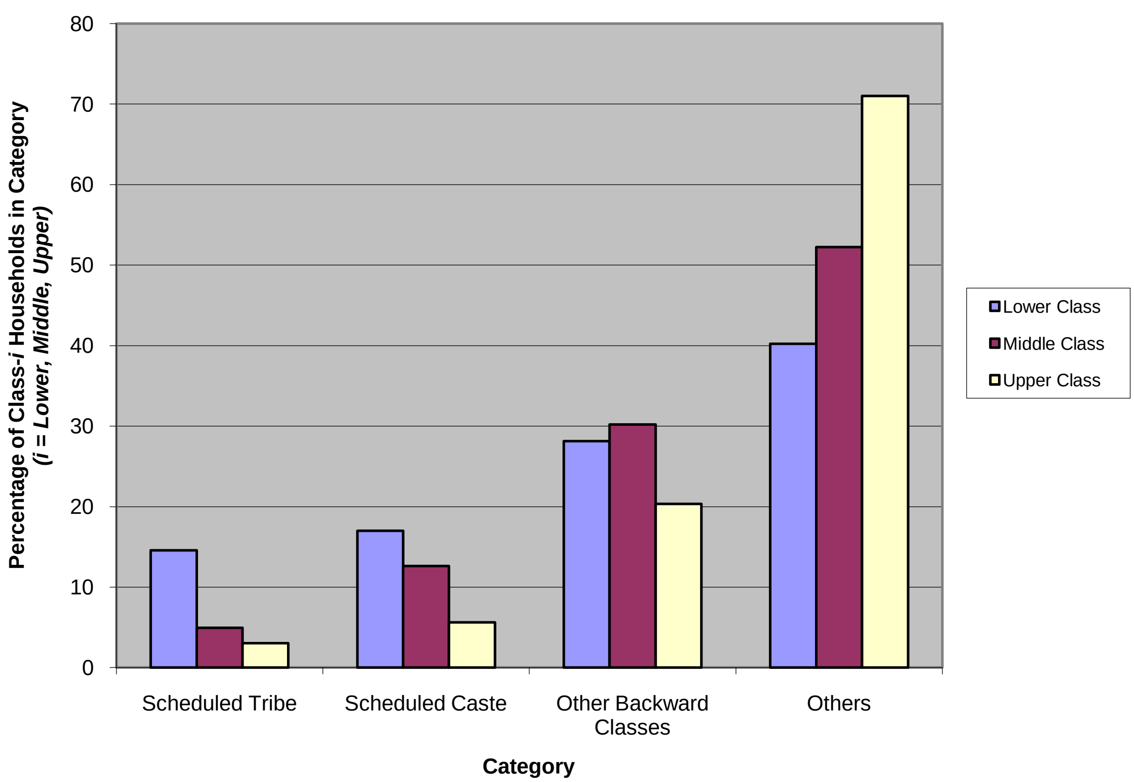
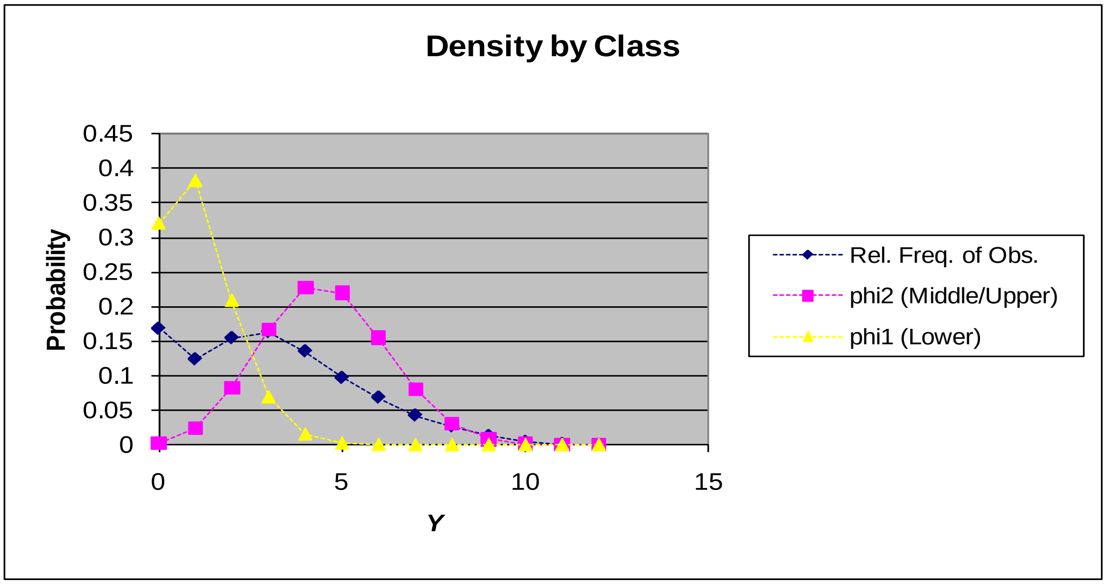
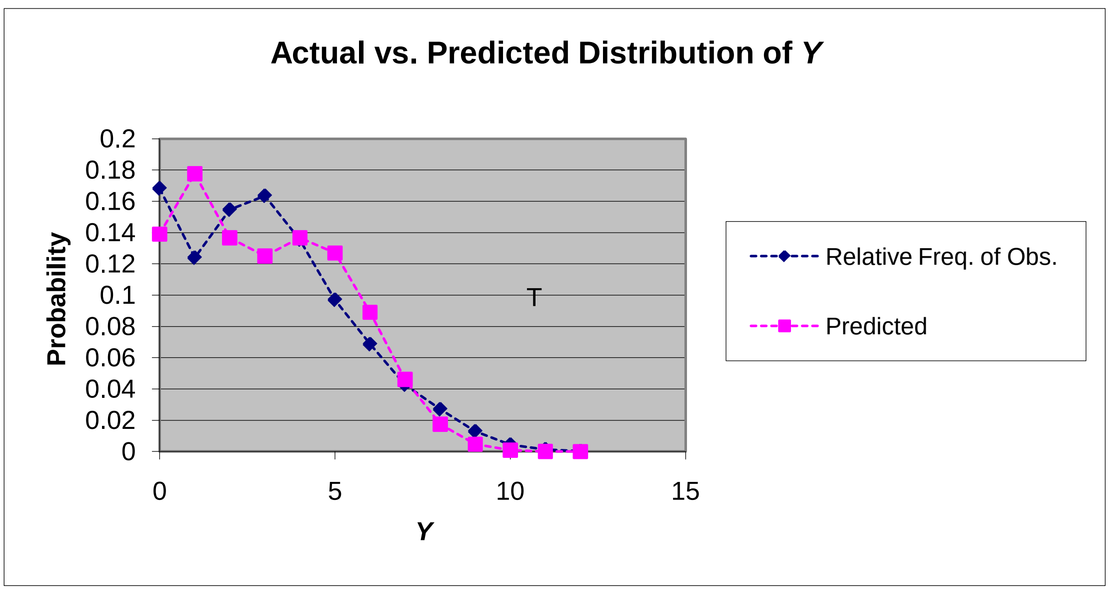


Fig. 12: Social Group, by Class



Appendix: EM Results for a Two-Component Mixture Model





Two-Components Model: EM Estimates

Class	Pop. Share	Prob. of Owning a Gd	Mean No. of Gds.
Lower	0.43	0.09	1.08
Middle/ Upper	0.57	0.38	4.52