Who are the Indian Middle Class? A Mixture Model of Class Membership Based on Durables Ownership[1, 2]

Sudeshna Maitra

Department of Economics,

York University

1038 Vari Hall

4700 Keele Street

Toronto, ON M3J 1P3

Email: smaitra@econ.yorku.ca

Phone: 416-736-2100 Ext. 77052

Fax: 416-736-5987

Original Version: July 2007

This Version: December 2007

WORK IN PROGRESS, COMMENTS WELCOME

[1] I am very grateful to Barry Smith, whose insights and encouragement have greatly benefited this research. All remaining errors in the paper are my own.

[2] The research was undertaken while I was Research Fellow at The Conference Board, New York, and I sincerely thank Bart van Ark for comments and support.

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

Sudeshna Maitra

York University

This Version: December 2007

ABSTRACT

The size and consumption habits of the Indian middle class have evoked considerable attention in the recent past. Yet the definition of the middle class has been nebulous at best. Most available estimates have used arbitrary income cutoffs to identify the classes, resulting in a wide range of estimates of the size of the middle class that depend on the researcher's assumptions about the appropriate cutoffs. I propose the use of a mixture model of class membership to identify and estimate the size of the lower, middle and upper classes in urban India, based on their distinct durables ownership densities. There are no external assumptions about who constitutes the classes, apart from the fact that their durables ownership patterns are *different*. Estimates using NSS data (55th Round, 1999-00) suggest that the urban middle class in India constitutes approximately 62% of urban households (which implies about 17% of all households) with mean ownership of 3 durable goods (out of 12). I also estimate the probability that each household in the sample belongs to a particular class and based on this information, back out some class-specific socioeconomic characteristics. The estimates suggest a larger urban middle class and lower class-defining income cutoffs than found (or used) in previous studies.

Keywords: middle class, durables ownership, mixture model, EM algorithm

JEL classifications: O15, I30, O10, O18

1. Introduction

Economists have long been interested in studying the middle class and for several reasons (see Banerjee and Duflo (2007)). Easterly (2001) argues that countries which have a larger middle class tend to have higher growth rates. Birdsall, Graham and Pettinato (2000) describe the middle class as the "backbone of the market economy and democracy in most advanced societies." One manner, it is argued, in which the middle class might accomplish the above is through their 'middle class values' – such as their 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 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, (see Murphy, Schleifer and Vishny (1989)) provide an impetus for rising income levels.

The size and characteristics of the Indian middle class have received attention for these 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 (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, India possesses a sixth of the world's population, and hence its middle class constitutes a significant portion of the global workforce as well as a substantial market for final products. Yet there have been surprisingly few attempts to define and identify the middle class in a rigorous manner. This paper seeks to address this gap in the literature by proposing a method to do so.

Prior studies of the middle class in India (Banerjee and Duflo (2007), Sridharan (2004), NCAER (2005), Ablett et al (2007), IBEF (2005)) have first imposed income (or expenditure) cutoffs for the different classes, and then proceeded to outline the characteristics (including consumption of durables) of the groups thus formed. Such an approach involves the use of several implicit assumptions – about who the different classes are and what their income levels must be – to which the results are extremely sensitive.

In this paper, I propose the use of a mixture model to model the distribution of durables ownership in urban India. The mixture model yields a class structure and membership probabilities which can be used to determine who constitutes the middle class. The appeal of a mixture model lies in its semi-parametric flavour. There are no external assumptions about who constitutes the classes, apart from the fact that the classes are *different*. I define the classes by an aspect of their consumption behaviour, viz. ownership of durable goods. The Expectations Maximization (EM) procedure then allows an estimation of the size and characteristics of the component classes in the population by identifying their *distinct* ownership patterns of durables. The unique solution generated by this approach provides an arguably more robust identification of the classes than has been obtained thus far.

The data comes from the 55th Round of the Indian NSS (1999-00). Durable ownership has featured prominently in discussions of living standards and the middle class (Banerjee and Duflo (2007), NCAER (2005), Ablett et al (2007), IBEF (2005)) in India. Hence I use data on durable ownership to define and identify the classes. I focus on the total of 12 durable items – 5 recreational goods (such as televisions), 4 household goods (such as refrigerators) and 3 transport goods (such as cars) – that a household may own at the time of interview. Since we are primarily interested in the middle class, which is largely perceived to be an urban phenomenon, I focus on the urban sub-sample of the NSS. However, the analysis may easily be extended to include the rural sub-sample as well.

I find lower, middle and upper class households to 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). The mean number of goods owned by households in these classes are, respectively, 0.3, 3 and 6.3. Small standard errors of estimates support the existence of three classes with distinct ownership patterns of durables.

The empirical approach involves maximum likelihood estimation. Maximum likelihood mixture models provide challenges in terms of parameter estimation and hypothesis testing. Here I use the Expectations Maximization (EM) algorithm for likelihood maximization (McLachlan and Krishnan (1996), Dempster et al (1977), Hastie et al (2001)). I provide a preview of the method in the next few paragraphs; Section 2 provides a detailed description of the model and methodology.

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 durable goods that a household owns at the time of interview. The objective is to estimate the population shares and durable-ownership density functions of the three component classes such that the likelihood of picking the sample is maximized.

The likelihood is maximized using the EM (Expectations Maximization) algorithm (McLachlan and Krishnan (1996), Dempster et al (1977), Hastie et al (2001)). The EM algorithm consists of 2 steps – the E step and the M step – which are iterated till convergence is obtained. Suppose that each household in the sample belongs to one of the three classes. Since actual class memberships are unknown, I estimate, for each household, the expected value of membership in each class conditional on the observed data on durable ownership. The conditional expectation is simply the probability that the household belongs to each class (since class membership can take values 0 or 1). This is the E ('Expectations') step of the algorithm. To perform this step, I begin with initial guesses for the parameters of the class-specific densities. The conditional expectation of class membership is substituted for the latent class membership in the likelihood function which is then maximized to obtain estimates of class shares in the population and the density parameters. This is the M ('Maximization') step of the EM. The E step is repeated with the parameter estimates obtained in the M step and the EM iteration continues till convergence is obtained. The likelihood of a sample based upon a mixture model is very complex and traditional numerical optimization techniques such as Newton-Raphson break down. The EM optimum coincides with the likelihood optimum but is reached (somewhat slowly) using iterated E and M steps.

How do the mixture model estimates compare with existing estimates of the Indian middle class? The mixture estimates suggest larger middle and upper classes than are found by Sridharan (2004), Ablett et al (2007) and the NCAER and IBEF 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 hard to compare for two reasons. First, Sridharan has followed the NCAER approach and defined the classes by arbitrarily setting income cutoffs. Second, each of his definitions of middle class includes the 'High' income category[1] 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

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

the middle class, and recommends the new method presented here for its intuitive approach to the issue. The per capita daily expenditure cutoffs used by Banerjee and Duflo (2007) to define the middle class are closer to those found in this paper, though still on the high side.

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 the EM 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

The data used in the 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 22 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 a household owns at least one piece of the durable 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, in combination, explain the overall density of Y observed in the sample.

There are two issues to be simultaneously resolved in applying a mixture model to the problem of identifying the classes. One issue concerns the choice of durable goods that may be expected to appropriately identify the classes. Another issue involves an assumption about the number of classes in the population, which is required in order to apply the appropriate mixture model. In response to the second issue, the appropriate number of classes is clearly 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, a better fit is obtained with three classes than with two (see Section 3.1 and Table C). Henceforth, I will refer to the classes as lower, middle and upper.

In response to the first issue of durable selection, note that our methodology will place households that own more durables in the higher classes. This follows from the fact that the variable Y used to define the classes is defined as the *sum* of durables owned by a household. Therefore, in order that the classes are correctly identified, the durable goods included in the basket should broadly satisfy three conditions. First, these must be non-necessary items, failing which even relatively poor households could be mistaken as upper class households due to the fact of their owning a large number of the goods. This precludes the inclusion of durable items such as furniture, bedding or luggage, in the relevant basket of goods. Second, the goods must be indicative of affluence. This broadly means that the goods are normal and more likely to be owned by affluent households, ensuring that higher values of Y indeed correspond to the higher (more affluent) classes. Third, there must be goods in the basket that identify households in different 'ranges' of affluence. For example, both normal as well as luxury goods would need to be included in the basket in order to correctly identify both middle as well as upper class households.

I focus on 12 (out of the 22 available) durable goods, that broadly follow the above conditions. These comprise 5 recreational goods (record player/gramophone, radio, television, VCR/VCP, tape/CD player), 4 household goods (electric fan, air conditioner, washing machine, refrigerator) and 3 transport goods (bicycle, motor bike/ scooter, motor car/ jeep)[2]. Figure B presents the proportion of households in each decile of monthly per capita expenditure, that own each of the 12 durable goods. With the exception of bicycles, all other goods are 'normal' for the entire range of per capita expenditure (PCE), in the sense that more affluent households are more likely to own them. Bicycle ownership increases up to the fourth decile but starts to fall thereafter. Banerjee and Duflo (2007) too report this pattern of ownership of bicycles in India. However, bicycles are included in my basket due to the third consideration listed above. In other words, since ownership of bicycles increases with PCE at low levels of PCE, it is likely to be a good identifier of middle class households vis-a-vis upper class households. Moreover, bicycles are also popularly perceived as a 'middle class good' in India (Banerjee and Duflo (2007), further justifying its inclusion in the relevant basket.

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

The bimodality and positive skewness of the distribution of Y in Figure A suggest that a mixture model may be an appropriate description of the underlying

[2] The 10 items that have been left out are items of household furniture/ furnishings, sewing machine, stove and pressure cooker/ pan. These are omitted on account of being necessary items or of not being very clearly indicative of affluence.

class structure. The objective is to identify the three distinct classes in the population such that their individual ownership densities or consumption patterns can, in combination, explain a distribution like that in Figure A. This idea, exploited in the Three-Component Mixture Model and estimated by an EM algorithm, is described in detail in the following subsections.

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 durable ownership of members. Assume that a household owns each good with a fixed probability (p_i), which depends on the class $(i = 1, 2 \text{ or } 3)$ to which it belongs. Assume also that each good is obtained independently. Hence the total number of goods owned by a class-i household follows a binomial distribution with parameters 12 and p_i[3].

[3] 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.

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 proportion of class i households in the population and $\phi_i(y;p_i)$ represents the (binomial) probability that the observation y comes from a class-i household. This is a Three-Component Mixture Model.

The likelihood function of the model described above 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.

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

Then the likelihood and log-likelihood functions may be written as

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

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

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, 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 $(\widehat{\gamma}_i^{(k)})$ of class membership, conditional on the data $(y_1, y_2, ..., y_N)$. Since class memberships are binary, $\widehat{\gamma}_i^{(k)}$ is also the estimated probability that a household belongs to class i, conditional on the data.

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

$i = 1, 2, 3.$

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.

$$\widehat{\pi}_i^{(k)} = \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij}^{(k)} \qquad (5)$$

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

$i = 1, 2, 3.$

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 proportion of class-i households in the population; $i = 1, 2, 3$

2. $\widehat{p}_i$: estimates of the probability with which a class-i household owns a durable good, $i = 1, 2, 3$

3. $\widehat{\gamma}_{ij}$: the probability with each each 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 class shares $\widehat{\pi}_i$ tells us the size of the urban middle class in India. Finally, the estimated 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 average per capita monthly expenditure, education of the household head, household type by employment and so on.

The next section presents the results.

3. Results

3.1. EM Estimates

The estimates produced by the EM algorithm are presented in Table 1 and Figures 1 to 3.

The numbers in column *(2)* of Table 1 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). 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 *(3)* 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 durable consumption behaviour[4].

An alternative interpretation of the numbers in Column *(3)* is that 52%, 25% and 3% of households in the upper, middle and lower classes, respectively, own a representative durable good. This interpretation allows an extrapolation of the size of the urban market for a representative durable good, as it specifies what proportion of the three classes will consume the good when it is introduced.

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

[4] 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) ($L \approx Lower$; $M \approx Middle$; $U \approx Upper$).

are reported in Column *(4)* of Table 1. The lower, middle and upper classes are found to own, on average, 0.3, 3 and 6.3 goods, respectively.

Figure 1 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, respectively. 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 2 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 Table C. The fit is clearly better in the Three-Component Model. Hence, three appears to be the minimum number of classes that provide a good fit to the data. This justifies the use of the Three-Component Model to describe the durable ownership habits of the urban Indian population.

Figure 3 plots the probabilities $\widehat{\gamma}_i$ that a household belongs to different classes i $(= 1, 2, 3)$. 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 certain to belong to the upper class (class 2).

3.2. Class Characteristics: A Descriptive Analysis

Using the mixture estimates of $\widehat{\pi}_i$ and $\phi_i(y; \widehat{p}_i)$ 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. As an example, suppose that there are 100 observations for $Y = 0$ and that the EM estimates predict that 60% of these belong to class 1, 10% to class 2 and 30% to class 3. I then randomly assign 60 of the 100 households with $Y = 0$ to class 1, 10 to class 2 and 30 to class 3. Likewise 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 are presented in Tables 2-3 and Figures 4-11 and discussed below.

Tables 2(a) – (b) and Figures 4(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[5]. Of these, electric fans and televisions are most popular among the top two classes, whereas fans and bicycles are

[5] 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 (3).

most popular among the lower class.

Table 3 reports the per capita monthly expenditures of households in each assigned class. These numbers suggest lower income cutoffs for the different classes than has been used in prior studies. 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 3), median annual household incomes are Rs. 41354.16 ($ 2840, PPP[6]), 58420 ($ 4013, PPP) and 104465 ($ 7176, PPP) for the lower, middle and upper classes respectively. The NCAER study 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[7]. Banerjee and Duflo (2007) define the middle class as having a daily per capita expenditure of $2 - $4 or $6 - $10. While closer than other studies to the middle class expenditures found here (the median daily per capita expenditure of the middle class is about $1.9 and the 99^{th} percentile is $8.5, see Table 3), these cutoffs are still on the higher side[8].

[6]Using a PPP conversion rate of Rs. 14.558 per US$ (WHO, PPP, 2000).

[7]The NCAER study divides households into 4 classes: Deprived, Aspirers, Middle Class and Rich.

[8]By Easterly's (2001) definition of the middle class (those lying between the 20th and 80th percentile of the consumption distribution), the PCE cutoffs for the middle class would be Rs. 490 and Rs. 1377. Using these cutoffs would exclude about 20% of upper middle class members

Figure 5 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. Despite the mean proportion of literate middle-class-household members being 77% (see Table 3), 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 6 presents a plot of household type by employment. Being urban residents, 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 7 which plots sources of household income. Over 50% of households in each class have reported income in the past year from

as per our estimates. Birdsall, Graham and Pettinato's (2000) definition (those lying between 75% and 125% of median income) would exclude at least a quarter of the middle class at each end of the distribution obtained here.

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 8 and 9 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 10 and 11 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, when classes are defined by their distinct patterns of durable ownership. 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 precisely estimated 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 magnitudes of the share estimates indicate a larger urban middle and upper class (62% and 18%, respectively) than were found in previous studies (Sridharan (2004), NCAER (2005), Ablett et al (2007), IBEF (2005)). However, these previous studies have relied on several 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 'the data to decide' who constitutes the three classes based on their distinct durable ownership patterns. The solution obtained is unique. This recommends the usage of mixture models to identify the classes and investigate the characteristics of component households.

References

[1] Ablett, Jonathan, Baijal, Aadarsh et al. "The Bird of Gold: The Rise of India's Consumer Market." McKinsey Global Institute Report, May 2007, (available at: http://www.mckinsey.com/mgi/publications/india_consumer_market/index.asp)

[2] Banerjee, Abhijit and Duflo, Esther. "What is Middle Class about the Middle Classes Around the World." Working Paper, December 2007

[3] Bardhan, Pranab. "The Political Economy of Development in India." Blackwell Publishing, 1984

[4] Birdsall, Nancy, Graham, Carol and Pettinato, Stefano. "Stuck In the Tunnel: Is Globalization Muddling the Middle Class?" Brookings Institution, Center on Social and Economic Dynamics WP No. 14, 2000

[5] Das, Gurcharan. "India's Growing Middle Class." *The Globalist*, November 5, 2001

[6] Deaton, Angus and Kozel, Valerie. "Data and Dogma: The Great Indian Poverty Debate." RPDS Working Paper, Princeton University, January 2005

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

[8] Easterly, William. "The Middle Class Consensus and Economic Development." *Journal of Economic Growth*, *6(4)*, 2001, pp. 317-335

[9] Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome. "Elements of Statistical Learning: Data Mining, Inference and Prediction." Springer Series in Statistics, 2001

[10] India Brand Equity Foundation (IBEF). "Economic Indicators: India's Middle Class Dream Takes Shape." 2005 (available at: http://www.ibef.org/artdisplay.aspx?cat_id=391&art_id=5788)

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

[12] McLachlan, Geoffrey J. and Krishnan, Thriyambakam. "The EM Algorithm and Extensions." Wiley Series in Probability and Statistics, 1996

[13] Murphy, Kevin M., Schleifer, Andrei and Vishny, Robert. "Industrialization and the Big Push." *Journal of Political Economy*, *97(5)*, 1989, pp. 1003-1026

[14] NCAER Report. "The Great Indian Market." 2005 (Preview slides available at: http://www.ncaer.org/downloads/PPT/TheGreatIndianMarket.pdf)

[15] Sridharan, E. "The Growth and Sectoral Composition of India's Middle Class: Its Impact on the Politics of Economic Liberalization." *India Review*, *3*(4), pp. 405-428, 2004

Table A: 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.

Fig. A: Distribution of Y

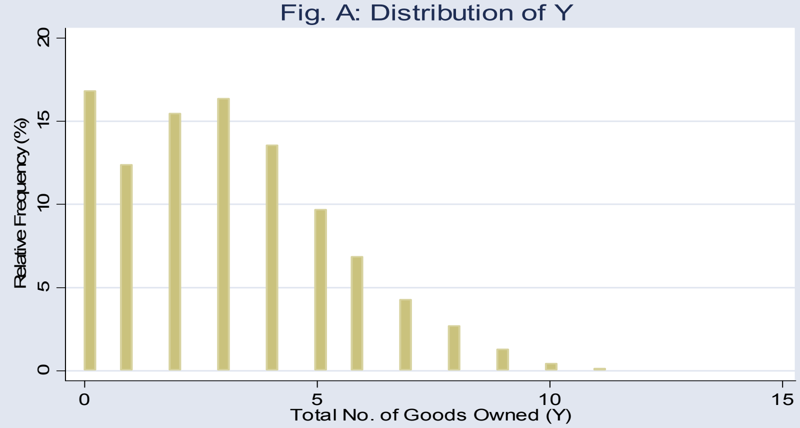


Fig. B: Durable Ownership by PCE Decile

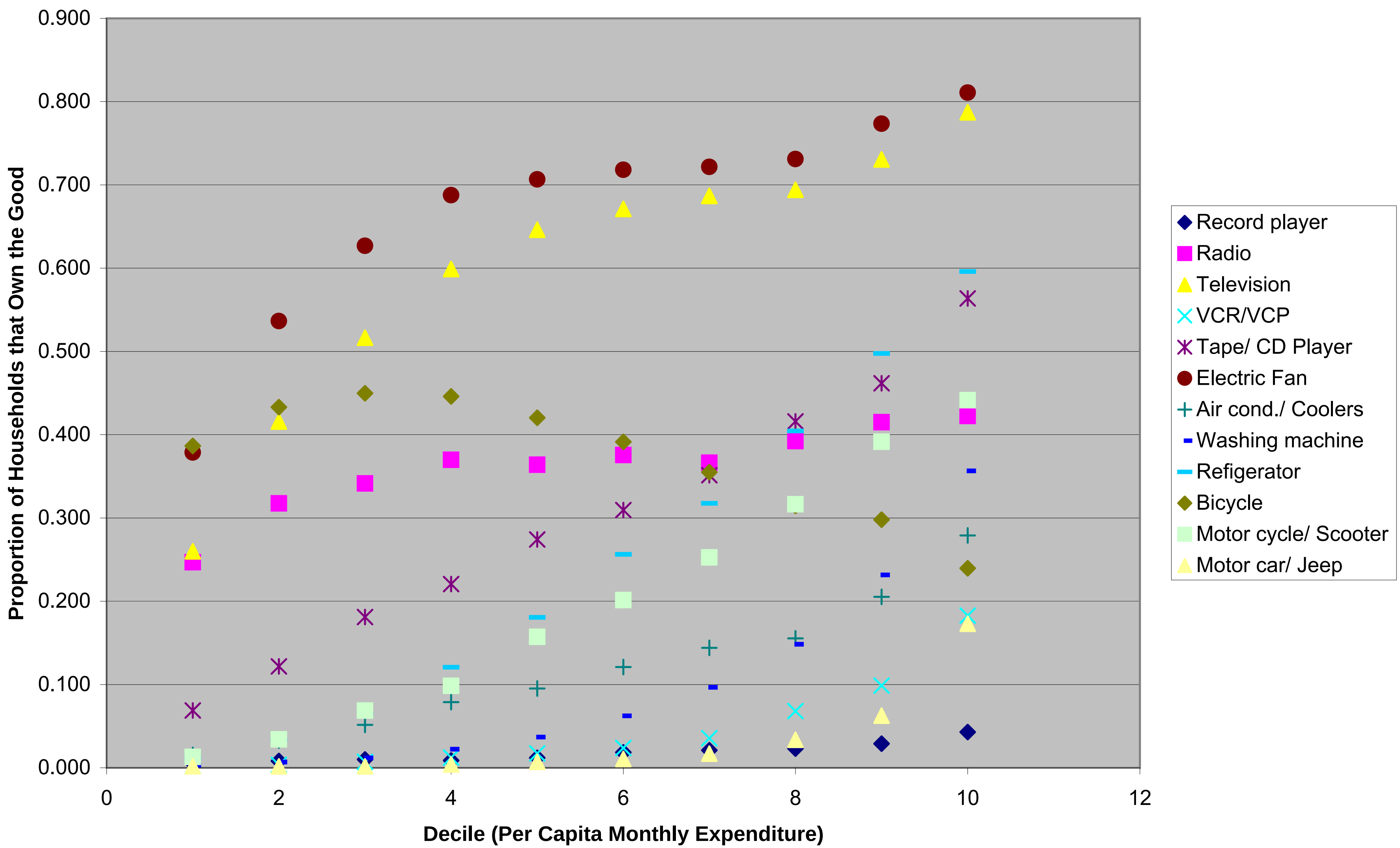


Table 1: Lower, Middle and Upper Classes in the Urban Sub-sample, Indian NSS, 55th Round (1999-00), N = 48,924 households

	EM Estimates (Std. Error)		
(1) Category (Class)	*(2)* Share of Urban Population	*(3)* Probability of Owning a Good	*(4)* Mean No. of Goods (of 12)*
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)

Table 2(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 2(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 3: Household Characteristics, by Class, in the Urban Sub-sample, NSS, 55th Round (1999-00)

Category (Class)	Per Capita Monthly Household Expenditure in Rupees [2000 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 [$54.35]	859.11 [$59.01]	17.00 [$1.17]	50528.00 [$3470.81]	423.00 [$29.06]	625.00 [$42.93]	981.00 [$67.39]	1421.00 [$97.61]	2791.43 [$191.75]	2.34	0.64	3.97
Middle (M)	961.79 [$66.07]	1772.39 [$121.75]	49.00 [$3.37]	205987.00 [$14,149.40]	532.00 [$36.54]	762.00 [$52.34]	1140.00 [$78.31]	1663.00 [$114.23]	3485.00 [$239.39]	2.38	0.77	4.65
Upper (U)	1469.57 [$100.95]	1109.97 [$76.24]	224.00 [$15.39]	35612.00 [$2446.22]	842.00 [$57.84]	1229.00 [$84.42]	1777.00 [$122.06]	2490.60 [$171.08]	5390.08 [$370.25]	2.41	0.88	5.12

Addendum: Percentiles of Per Capita Monthly Expenditure (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: PPP Conversion Rate: Rs. 14.558/US$ (WHO, PPP, 2000)

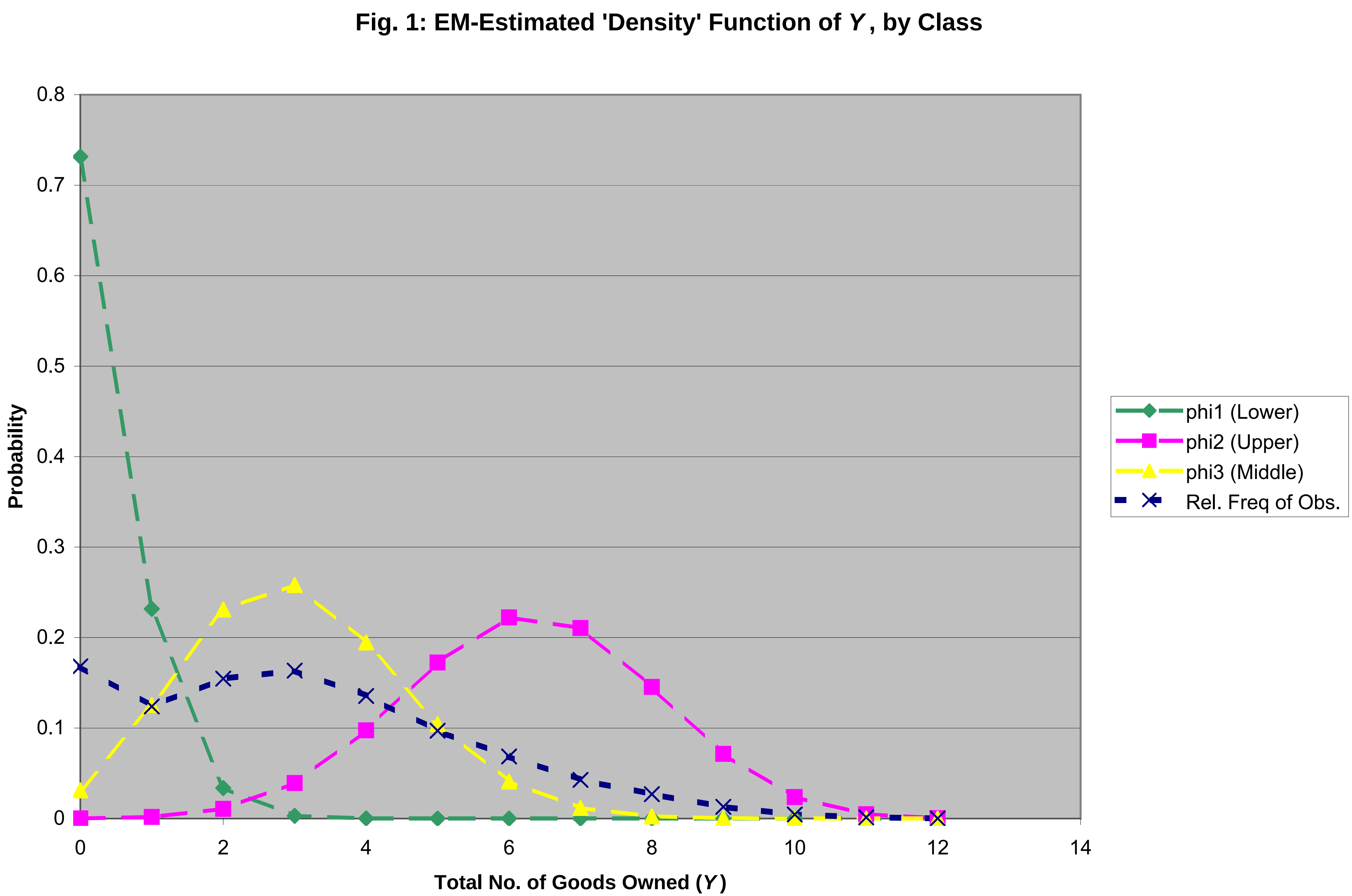


Fig. 1: EM-Estimated 'Density' Function of Y, by Class
Probability
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
Total No. of Goods Owned (Y)
phi1 (Lower)
phi2 (Upper)
phi3 (Middle)
Rel. Freq of Obs.

Fig. 2: Actual vs. EM-Predicted Distribution of *Y*

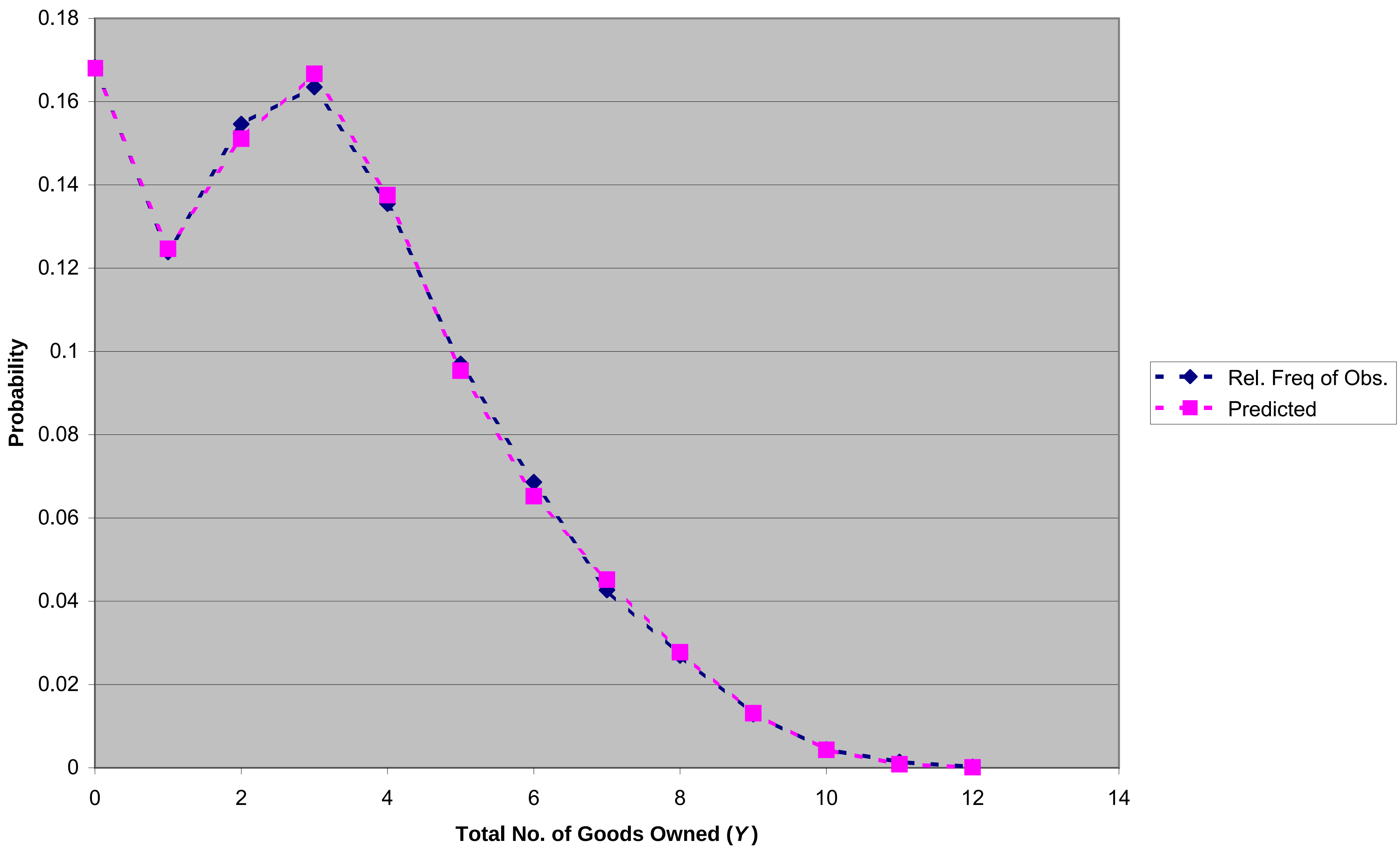


Fig. 3: EM-Estimated Probability of Belonging to Each Class

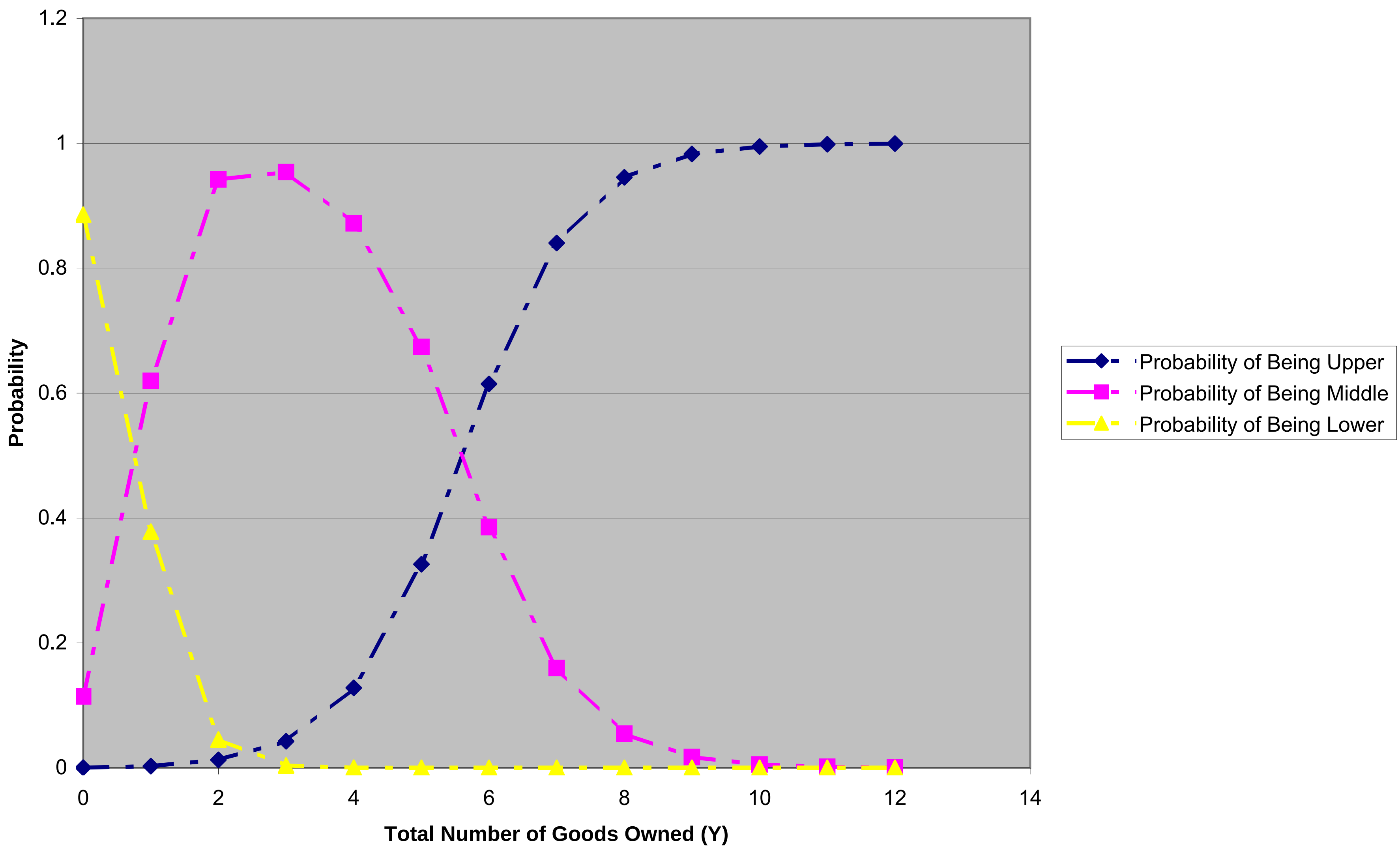


Fig. 4(a): Ownership by Durable Categories by Class, Urban Sub-sample, NSS 1999-00

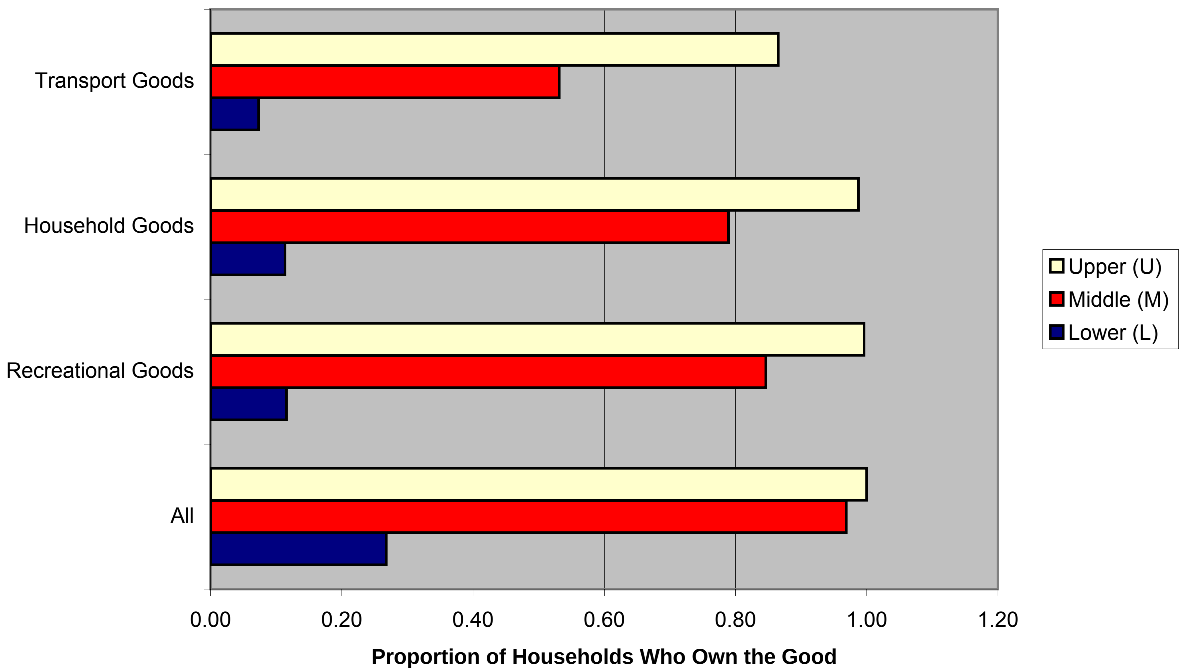


Fig. 4(b): Ownership of Individual Goods by Class, Urban Sub-sample, NSS 1999-00

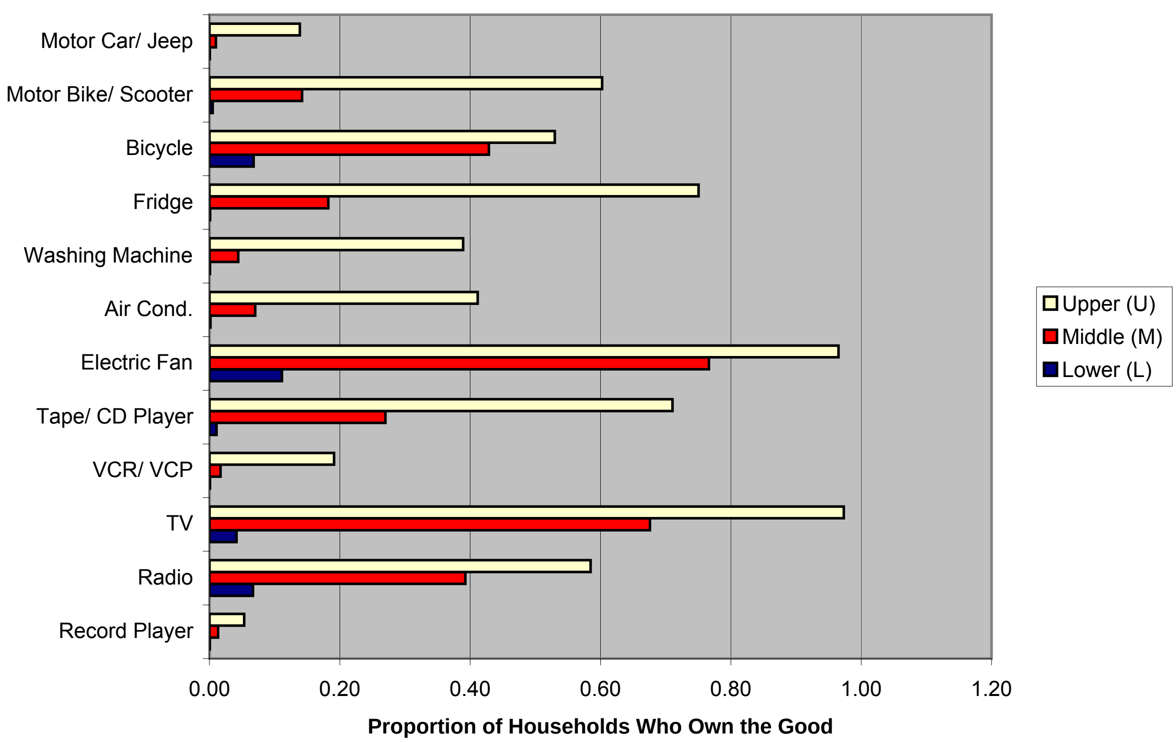


Fig. 5: Level of Education, by Class

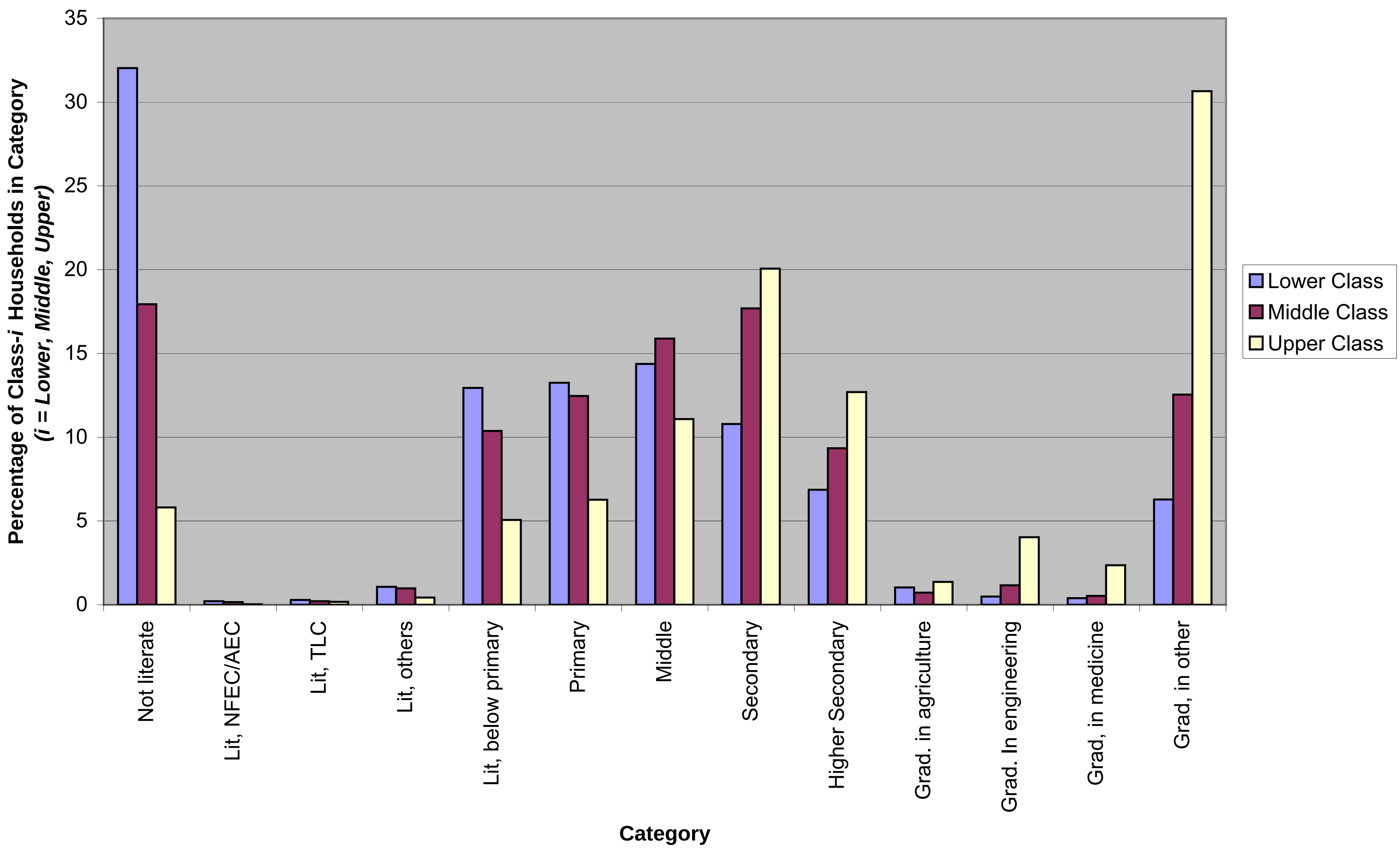


Fig. 6: Type of Employment, by Class

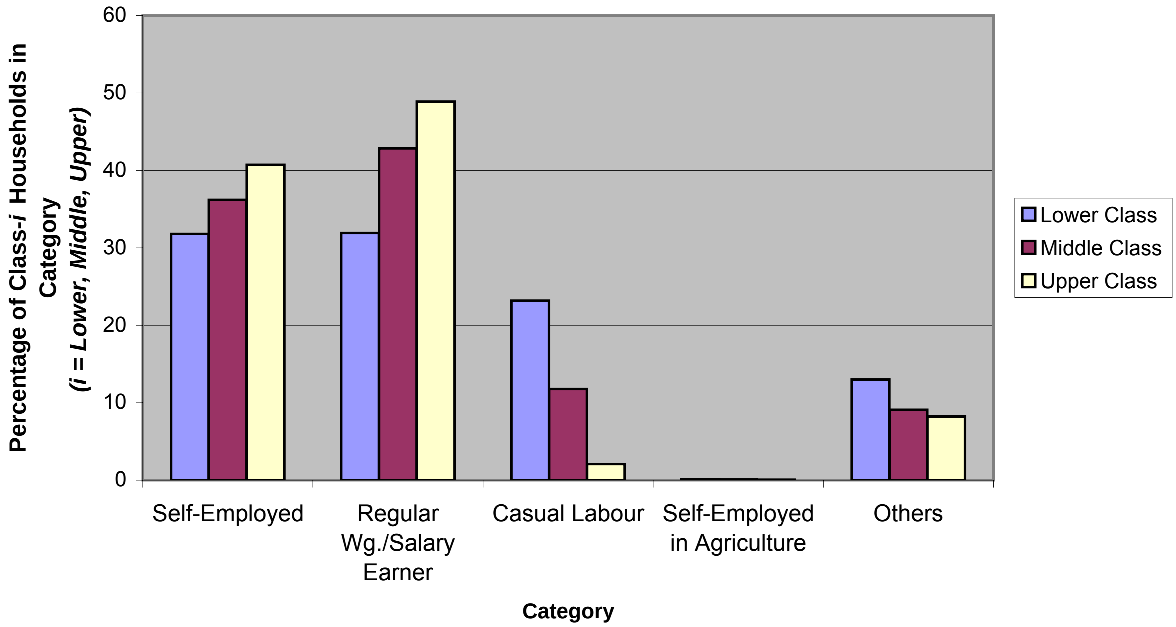


Fig. 7: Land Ownership & Source of Income, by Class

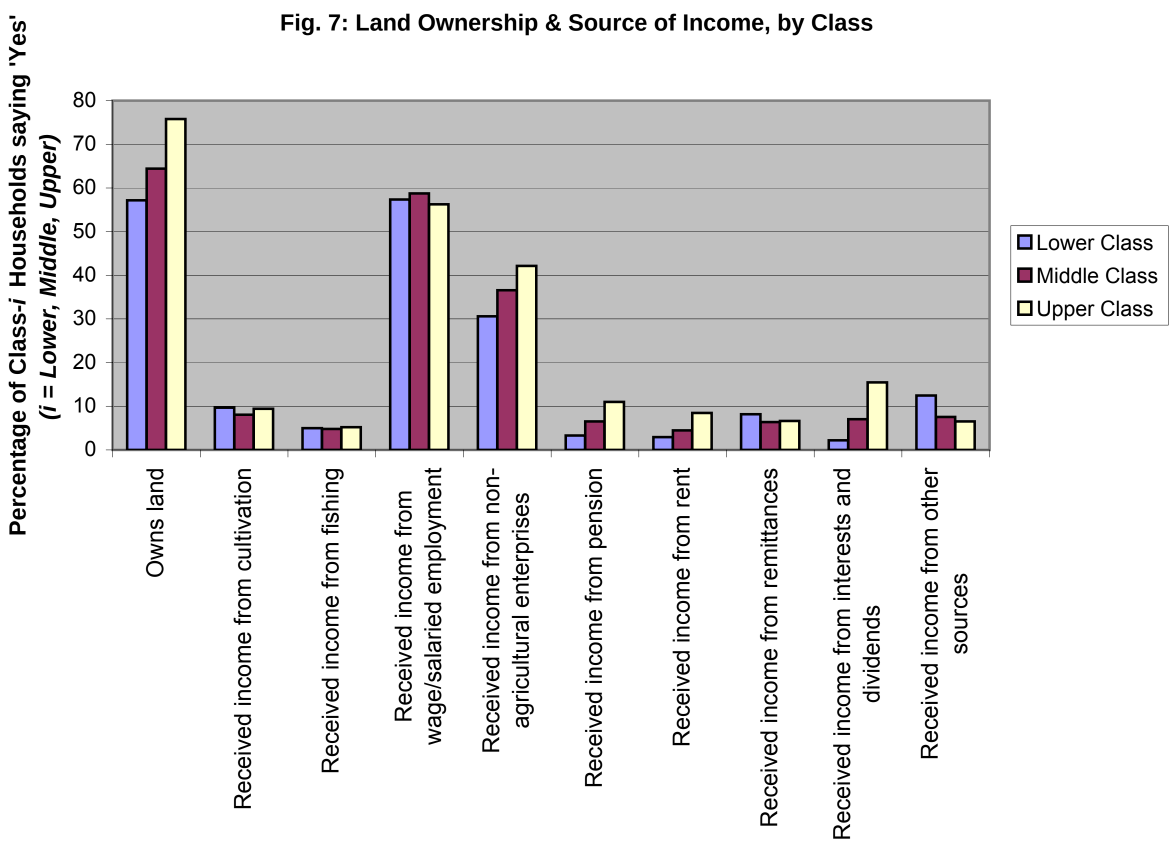


Fig. 8: Primary Source of Energy Used for Cooking, by Class

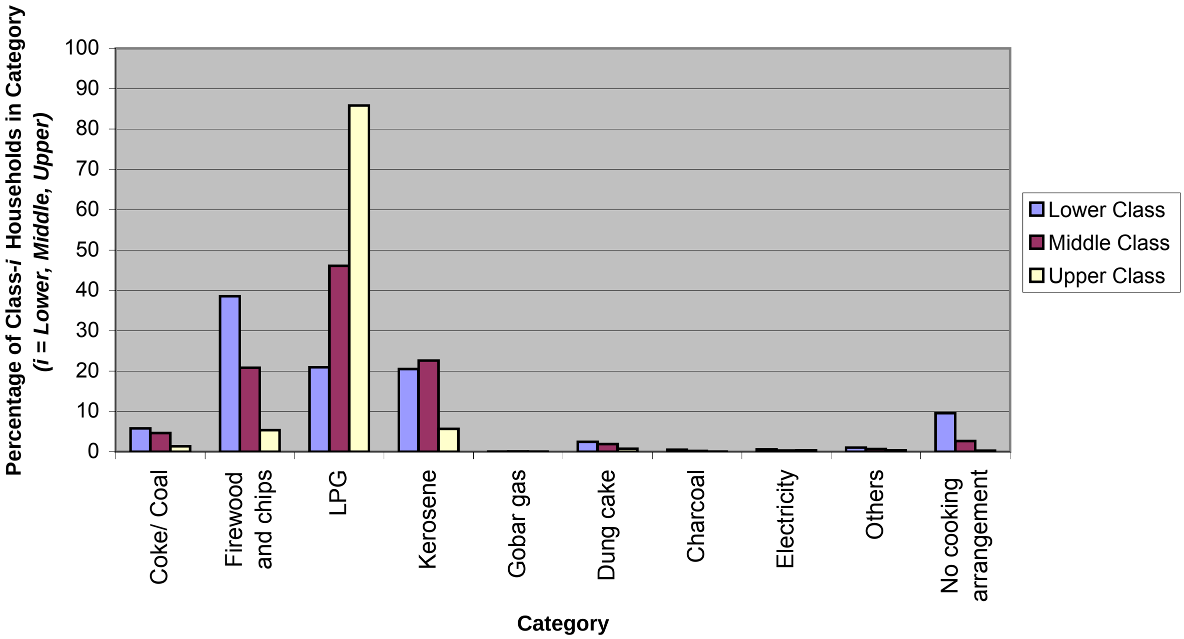


Fig. 9: Primary Source of Energy Used for Lighting, by Class

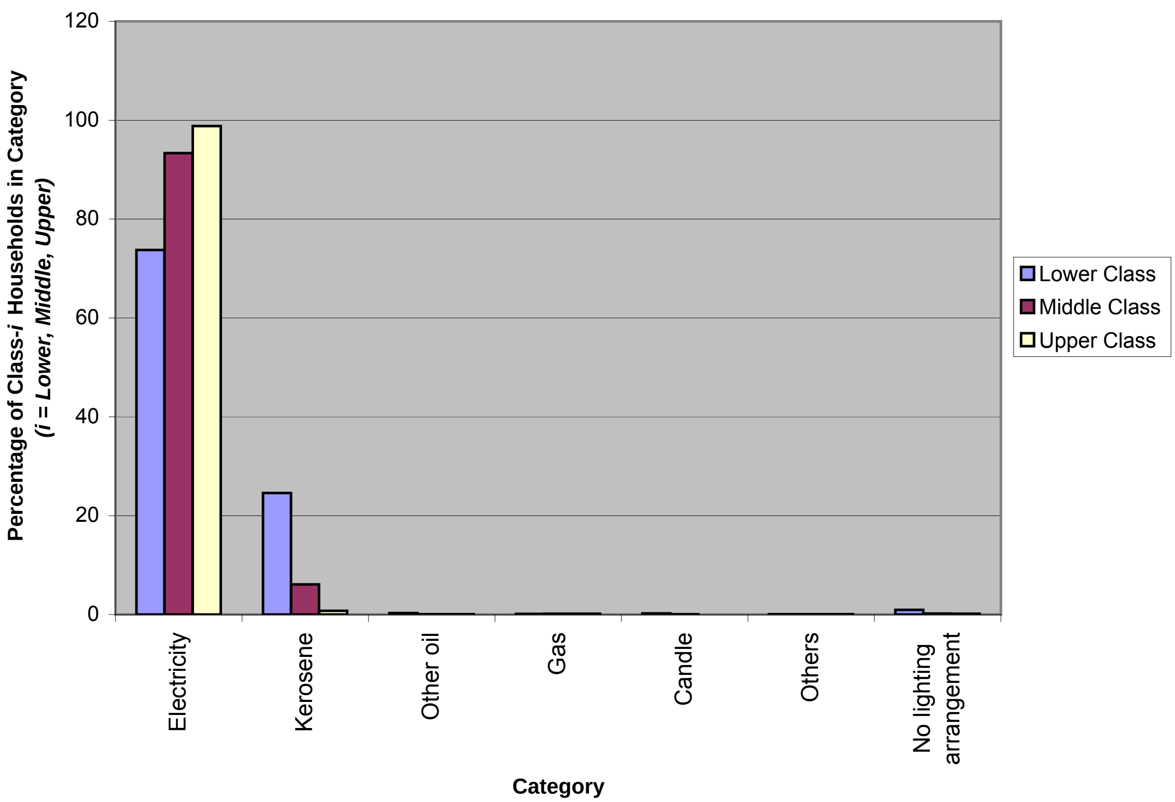


Fig. 10: Religion, by Class

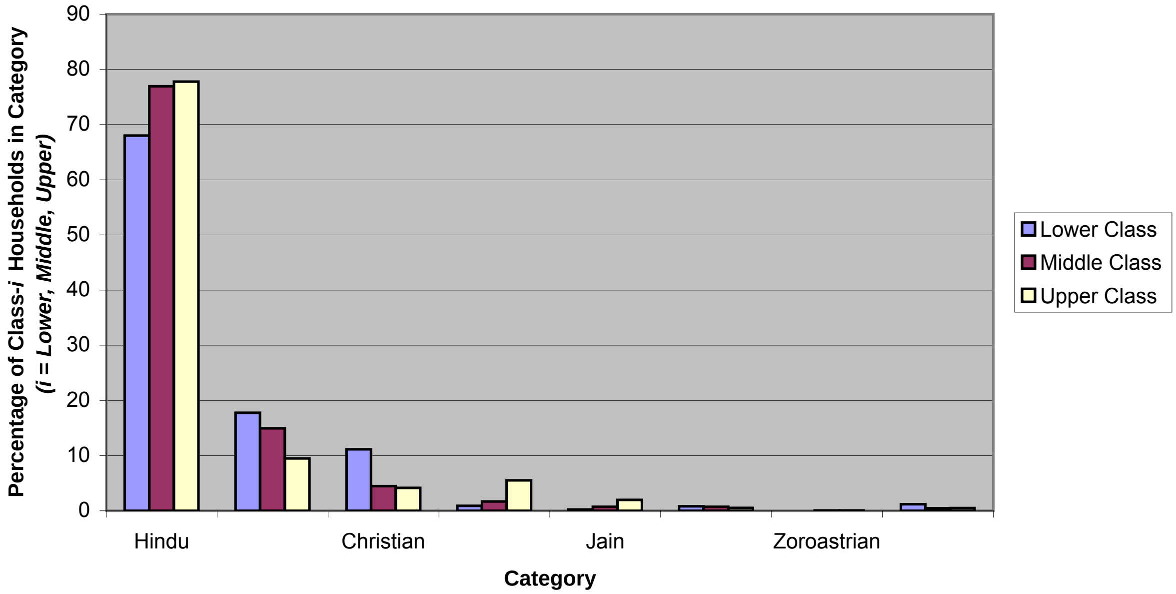


Fig. 11: Social Group, by Class

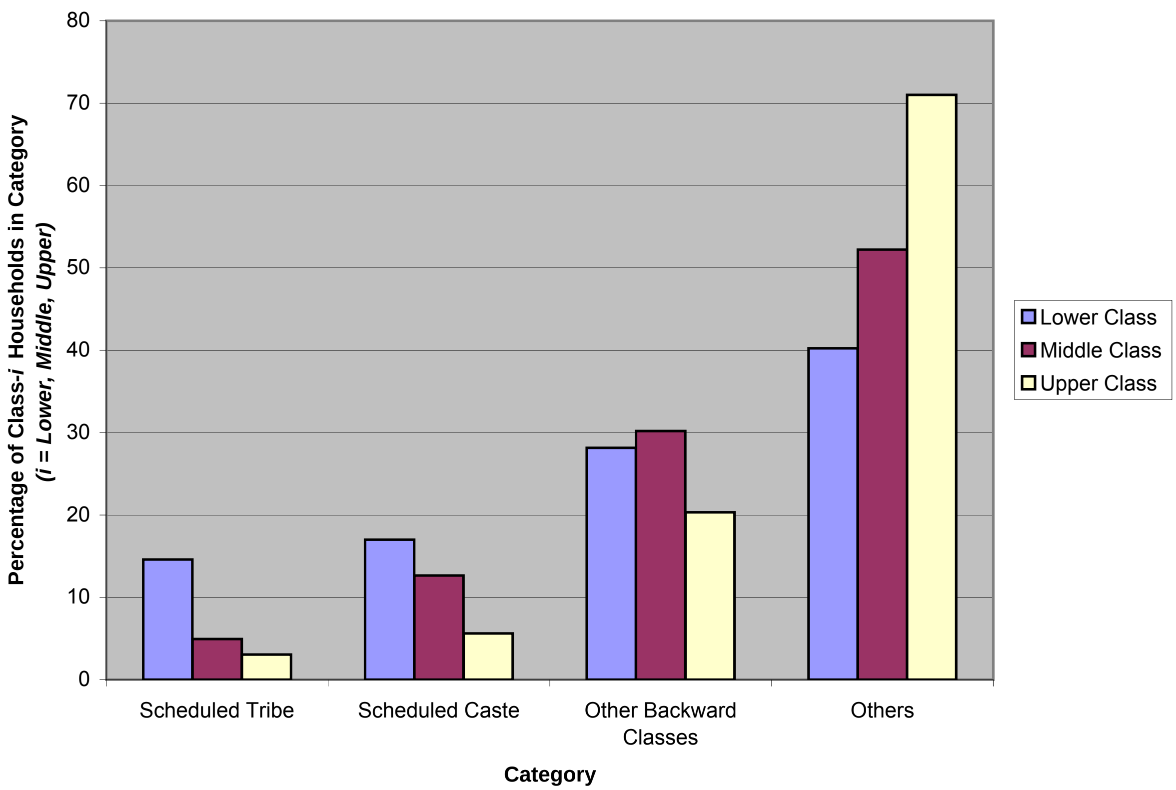
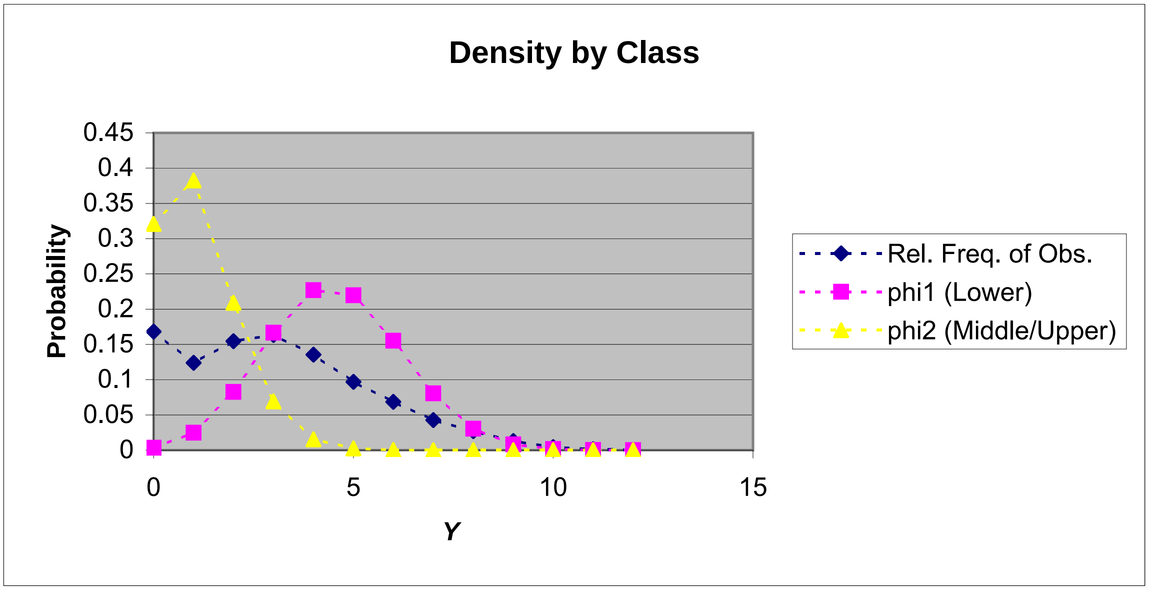
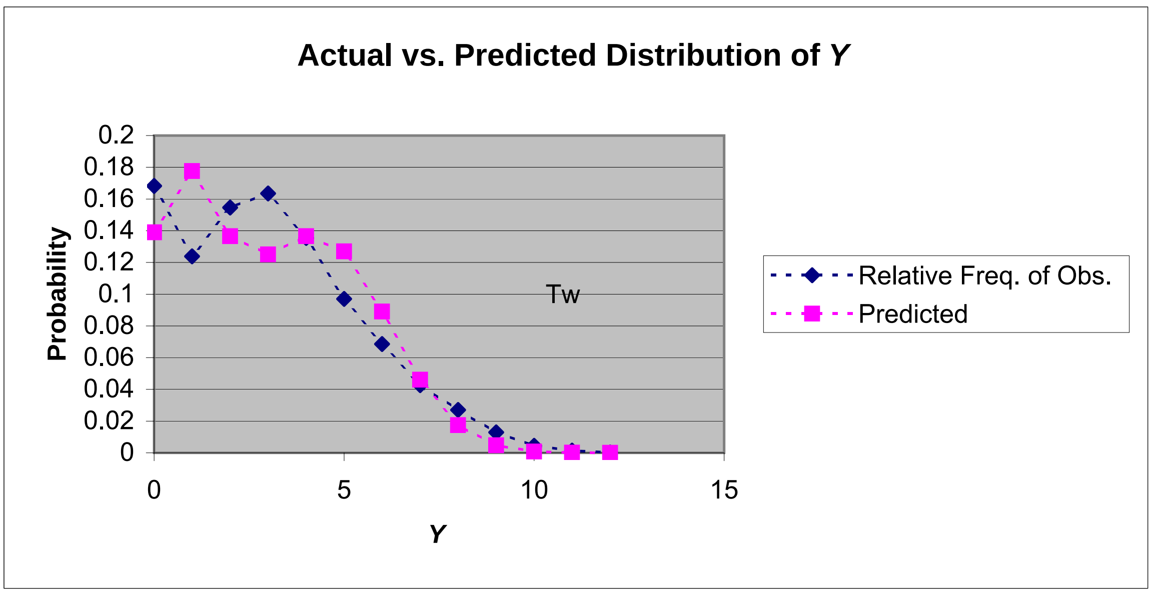


Table C: EM Results for a Two-Component Mixture Model





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