Identification and characterization of middle class in India and its comparison with other economic classes*

R Ahalya †
and
Sourabh Bikas Paul ‡

August, 2017

Abstract

In this paper, we attempt to identify middle class in rural and urban India respectively and compare its characteristics with the lower class and the upper class. The middle class is an ambiguous category and has not been explored much in the discipline of economics. We identify economic classes so as to study how economic status is related to various social outcomes. For this purpose, two approaches are used: the cut off approach and the probabilistic approach. In the cut-off approach, pre defined monthly per capita consumption expenditure(MPCE) cut offs are used to define the poor, the lower class, the middle class, the global middle class and the rich. In the probabilistic approach, mixtures model is used to cluster the households as lower class, middle class and upper class. The results so obtained through both these approaches are compared so as to be studied and extended further.

*

†Indian Institute of Technology Delhi, New Delhi. Email: ahalyaramanathan@gmail.com
‡Indian Institute of Technology Delhi, New Delhi.

1 Introduction

The middle class in India is a largely ignored group as far as academic research in economics is concerned although their presence is widely acknowledged. The middle class is considered as the most important class for driving economic growth and providing a huge market for consumer durables in a country. They are considered as a large and powerful group capable of bringing social and economic reforms. However, the middle class cannot be studied in isolation but only in comparison to the poor at the lower end and the rich at the upper end. Broadly, those people who have enough cushioning against any sudden economic crisis, either at the household or national level can be classified as middle class (Ravallion, 2010). This means a middle class household must have a secure stream of income or assets in order to overcome any drastic financial crunch. But economic security alone is not what makes a household belong to middle class. There are other social and cultural aspects characteristic of the middle class. The middle class are supposed to possess at least some benchmark level of human development standards. For example, the parents of children belonging to a middle class family possess at least some minimum educational attainment which entitles the child with access to better education and social networks for further upward mobility. Similarly, a middle class household has some basic household amenities necessary for enjoying a reasonably comfortable standard of living (Thurlow et al., 2015; Banerjee and Duflo, 2008). Apart from this, the middle class possess certain values that encourage them to give up conspicuous present consumption so as to ensure future financial security for themselves and their future generations. Unlike the rich,however, they are not so well endowned that they do not have to earn a living. Thus, the middle class have a greater drive to save and invest compared to the poor and the rich classes due to their aspirations for further upward mobility (Birdsall et al., 2000). This is reflected in the choices beyond the economic sphere i.e. the level of social, cultural and human development that they choose to acquire given their income levels. That is why it is necessary that the middle class are not just defined as an economic group but as a socio cultural group possessing certain distinct attributes (Pierre, 1979). In this paper, however, I shall be using only economic criteria to identify

the middle class and then observe how it behaves with respect to these socio cultural factors in comparison to the other economic classes.

Despite a few studies regarding the Indian middle class, there is still no clarity regarding the criteria to classify the middle class in India. In a globalized world, it is difficult to estimate what comprises a comfortable standard of living. The idea here is that the tripping line indicating the lower limit for entering the middle class must be such that at that level of income, the individual has sufficient cushioning against falling down to the lower class subsistence level of living. The question here is what all factors are to be taken into account while determining this tripping line (Lopez-Calva and Ortiz-Juarez, 2014). For instance, apart from the level of income, possession of some property, level of job security and educational and skill attainment of an individual may determine his/her vulnerability to fall to subsistence level. Another thing that needs enquiry is the nature and extent of disparities within the spectrum of the middle class and the basic requirements for an individual so as to enable upward movement along the scale of middle class. This can be determined by the differences in the wealth, parental occupation, level and quality of education and the availability of social networks for an individual (Banerjee and Duflo, 2008). Finally, another interesting feature observed lately among the middle classes is the changes in their social behaviour in line with the process of westernization (Fernandes, 2000). Along with this social change, the tastes and preferences of the middle class are changing, especially among the upper layers such that the benchmark for a comfortable standard of living has been moving upwards. This feature poses the question of how the pressure to keep up with the Joneses is pushing the middle class to strive to move up the socio economic ladder. In practice, it is difficult to incorporate all these factors into a measure of economic class due to constraints in the availability of data .Therefore, the approches used in the paper do not intend to find foolproof measures of middle class, rather include available dimensions that may be reflective of what all constitute a measure of middle class.

It is projected that the proportion of middle class in the total population will rise to 41 % by 2025 according to a report by Beinhocker et al. (2007). These

rising middle classes are to be the drivers of future economic growth and provide a huge market for consumer goods creating greater avenues for investment. But given the huge diversity in the perception of who are middle class, this may not be completely true. On one hand, there is an increasingly vulnerable lower middle class who lack of proper insurance combined with increasing job insecurity which may hinder their upward mobility. On the other hand, the rising aspirations and standard of living among the upper middle class(or the new global middle class),are likely to make them a part of the newly rich but this may again be a very small section of the Indian population. Therefore, this story of the Indian middle class seems to need a relook so as to determine what constitutes the middle class in India and how diverse a group of people are included within this class. This may have important policy implications in order to create a stable middle class to launch the country on the path of growth and equity.

2 Literature review

The focus of this paper is to identify and characterize the economic middle class in India. The other economic classes will be looked at in comparison to this class. Therefore, we shall be discussing studies conducted in regard to identification of middle class in India and elsewhere. They have used varied standards and methodologies to quantify the presence of a middle class. The methods used in this paper have been attempted to be an improvisation on these existing approaches. We briefly discuss some literature related to poverty measurement and identification of the rich and the controveries associated with these studies that may be helpful in highlighting the pros and cons of the approaches used in the paper at a later stage. Finally, we shall be briefly discussing some papers that helped us understand the application of the mixtures model and related statistical techniques employed in this paper.

NCAER report (2010) has made an attempt to define the criteria for middle class in the Indian scenario. NCAER has divided the middle class into two categories :the seekers with annual income between Rs.200,000 and Rs.500,000 and strivers with an annual income between Rs.500,000 and Rs.1000,000. Meyer

and Birdsall (2012) use 66th round NSS consumption expenditure data to find that only 5.88% of the total population of India are middle class using the NCAER criteria of identifying middle class. According to Fernandes (2000), there has been an ongoing drastic change in the socio economic characteristics of the newly emerging middle class since the liberalization of 1991. While the general picture that emerges of the urban middle class post liberalisation is that of an increasingly consumerist and upwardly mobile segment but the job security enjoyed by the pre-liberalization is no longer there. Also, the distinctions between the different layers of the middle class has been widening. Ravallion (2010), defines middle class as all those individuals whose per capita consumption expenditure is between $2 and $13 per day in developing countries. According to Banerjee and Duflo (2008), classified all those with a per capita income between $2 and $10 per day as middle class.

The above studies attempt to identify middle class in India on the basis of an income/consumption expenditure range. Maitra (2007) has attempted to create a mixture model of durables ownership so as to identify the middle class. She has made no ex ante assumptions about who constitute the middle class and has classified them solely on the basis of their durables ownership. Using this approach on 55th round of NSS consumption expenditure data, she finds that 62% of the urban households in India were middle class. This estimate of middle class is close to Ravallions estimate of middle class as 56% of the households. According to Maitra, this method scores over the other methods to identify middle class used currently since possession of durable goods is a better indicator of the permanent income of households.

In the context of America, the Pew research centre has conducted the most recent study of middle class in 2014 (Pew Research Center, 2012).The criteria for middle class used in case of America is that all those households whose income is between two thirds to double the median income of the United States can be considered as middle class. The study has found that the American middle class has been shrinking over time and has become poorer over time. Nationally, the median income of the American middle class households declined from $77,898 in 1999 to $72,919 in 2014, which is a loss of 6%. Also, 203 of the 229 metropolitan

areas surveyed had shown a decline in the share of middle income households. According to Kharas (2010), the Chinese middle class is likely to expand rapidly in the next few decades and has the potential to replace the United States as the consumer of last resort, given the declining middle class share in America.

Many other studies have used various other standards to define middle class globally. According to a World Bank report (Kharas, 2011), the minimum threshold income for a middle class person is $ 10 per day and the maximum income is $50 per day. Easterly (2001) categorizes all those individuals between 20th and the 80th percentile of the overall consumption expenditure as middle class. According to Birdsall (2010), those individuals with a per capita consumption between $10 per day and the 95th percentile of the overall consumption expenditure (specific to each country) are to be classified as middle class. This gives a common absolute minimum threshold for all countries and a relative maximum threshold specific to each country. According to Milanovic and Yitzhaki (2002), defined those living between the mean per capita incomes of Brazil and Italy can be considered as middle class.

Another interesting work that attempts to define Africa's middle class (Thurlow et al., 2015) uses three criteria to categorize the middle class that are considered as jointly necessary and sufficient for a household to shield itself against economic vulnerability while providing scope for upward mobility. The three criteria used by the authors are decent housing amenities, secondary school completion and skilled employment. The housing amenities in turn comprise of availability of piped water, a flush toilet and electricity from main grid, solar panel or private generator. The criteria for fixed for education is secondary school employment and the criteria for skilled employment is non agricultural and non elementary occupation or own business. They show that a large fraction of the households that are considered as middle class according to the consumption expenditure ranges(as given by African Development Bank) fail to satisfy these criteria and there is great variation across African countries.

In addition, there have been innumerable studies and debates regarding the measurement of poverty in India. The Rangarajan committee report (2014) fixed the poverty line at a daily per capita consumption expenditure of Rs.32

in rural areas and Rs.47 in urban areas in 2011. This was an upgradation over the earlier norm which was based solely on calorie intake while this poverty line was fixed taking into account expenditure on health, education and other necessary items. According to Patnaik (2010), upgrading the poverty line from time to time by using the price index as a deflator is inappropriate since the price index is not an adequate indicator of rise in cost of living, which is much higher than that indicated by this measure. Swaminathan (2010) is critical of the earlier poverty line fixed by the Tendulkar committee report (2009) on the grounds that it is inadequate in providing for minimum nutritional, health and educational outcomes.On the other end of the spectrum, there have been some studies to identify the rich in India. A notable study of the rich in India has been done by Banerjee and Piketty (2005) who made use of income tax returns data for the purpose. According to them, consumer expenditure surveys like that conducted by NSSO are known to underreport consumption by the rich, mostly because the rich do not report their incomes accurately or do not cooperate with the survey personnel. They have found that the rich have been able to increase their share of income since the liberalization of the economy indicating that they were the ones who appropriated the greatest share of the fruits of liberalization.

Dempster et al. (1977)put forward a broadly applicable algorithm for computing Maximum Likelihood estimates from incomplete data known as the Expectation Maximization Algorithm. McLachlan and Jones (1988) applied the Expectation Maximization algorithm for fitting finite mixture models for data which are available only in grouped form and may be truncated. Chu Chan(2016) demonstrates in his blog how to fit a mixture model in R using the Expectation Maximization algorithm. Here, he applies the Gaussian mixtures model to the Geyser dataset that is assumed to be normally distributed. Wood (1999) uses a binomial mixture model for the geometric estimation of the mixing distribution. These are the few papers that provide the framework for the Expectation Maximization algorithm used for clustering datasets in this paper.

3 Data

We have used the 68th round of NSSO consumption expenditure data to identify middle class in India. Covering the period 2011-12, it gives estimates of the Monthly Per Capita Expenditure (MPCE) at the household level and its commodity wise disaggregation apart from a few other socio economic characteristics of the persons in the households. The data comprises a sample of 101,662 households and 464,862 individuals. It comprises of 41,964 urban households and 59,681 rural households. Depending on the reference period used, data for three different types of MPCE have been given. We use MPCE data with Uniform Reference Period (URP) for the purpose of my study, the reason being it is uniform across time and is useful for comparison over time. NSS data with its large sample size covering almost all Indian states, is the best available data source for measuring living standards. No other data source provides such extensive coverage on consumption, social class, education, durable ownership etc. that are relevant here. There are, however, minor problems with the data base, like the presence of outliers and missing values. So, while working with such data, we had to remove the outliers and missing values to obtain reliable results. Also, a small part of the data may have been lost during the process of merging different levels of the dataset, but that does not affect the results much.

4 Methodology

In this paper, two distinct approaches are used to group households into various economic classes so as to identify the middle class. The first one makes use of pre-determined cut offs of per capita consumption expenditure for this purpose while the second approach is based on the Expectation Maximization algorithm that assigns classes to the households with a probability rather than certainty. Both these approaches along with the underlying methodology is explained as follows :

4.1 Consumption expenditure range

Since consumption expenditure data has been used, it is not found to be appropriate to use the income criteria since consumption expenditure is far from an exact proxy for income. Therefore, for the purpose of my analysis, we start with the criteria used by Banerjee and Duflo (2008) to categorize middle class. They have used two different sub ranges within their $2 to $10 range to distinguish between the lower and upper end of the middle class i.e. they have classified those with per capita income between $2 and $4 per day as lower middle class and those with per capita income between $6 and $10 as upper middle class. At 2011 PPP exchange rate of about Rs.15/$, the lower end of this range is Rs.30 per day per person. But this amount is too low considering that the Rangarajan Committee report fixed the poverty line in 2011 as Rs.32 and Rs. 47 per day per person for rural and urban areas respectively. The middle class are not just the individuals above the poverty line but are supposed to have a high enough income so that they are not too vulnerable to fall below the poverty line. Therefore, the lower limit for entering middle class is fixed as Rs.73 per day for rural areas and Rs.98 per day for urban areas which is approximately double the poverty line for rural and urban areas respectively.This cut off is obtained by adding the average per capita expenditure of households on health, education and house rent(imputed) to the poverty line for rural and urban households respectively. Given that each month would consist of 30 days on an average, this would be Rs.2190 per month in rural areas and Rs.2940 in urban areas. The households with MPCE above poverty line but below the cut off set for middle class are considered to be vulnerable to poverty. The poor and the vulnerable households are jointly considered as lower class. The upper limit of MPCE for a household to be considered as middle is fixed as the $95th$ percentile of the MPCE range for rural and urban households. This is same as the relative upper cut off used in Birdsall (2010). The criteria set here for middle class households is appropriate for being classified as middle class according to Indian living standards. For being classified as middle class by global standards, Kharas (2011) gave the criteria that those households with a daily per capita consumption expenditure between $10 and $100 are global middle class. So, all those households with

a daily per capita consumption expenditure above $100 can be considered rich by global standards. In this way, we classify households as poor, vulnerable, middle class,global middle class and rich using their per capita consumption expenditure.

4.2 Mixtures model

The mixtures model is used to analyze complex distributions comprising of both observed and latent variables. When a joint distribution of observed and latent variables is defined, the distribution of the observed distribution alone can be obatined through the process of marginalization. Through this process, the joint distribution can be interpreted in terms of the latent variables.In the present study, we use mixtures model as a tool to cluster data. Firstly, we use a non-probabilistic method called K-means clustering using some initially defined parameters for the distribution, in order to cluster the data set into the required number of components. Then, we use the priors so obtained to introduce the discrete latent variables, so as to assign the data points to specific components of the mixture. For this purpose, the technique generally used is that of EM-algorithm mixtures models, especially gaussian mixture models are widely used in data mining, pattern recognition, machine learning and statistical analysis. In this paper, we employ two types of mixture models: the binomial mixtures model and the gaussian mixtures model. We will henceforth be discussing these two models in detail. Before that, however, we will be discussing the method of K-means clustering in some detail.

K-means clustering: Consider a data set $(x_1, ..., x_n)$ consisting of n observations of a random variable x. The prototype associated with each cluster k is μ_k which represents the mean or the center of the cluster. The aim here is to find an assignment of data points to clusters as well as to respective μ_k values, such that the sum of squares of the distances of each data point to its closest vector μ_k, is a minimum. For each data point, a corresponding binary variable $r_{nk} \in (0, 1)$ is assigned which takes value 1 if the data point belongs to cluster k, otherwise zero. The objective function to be maximized here, known as a

distortion measure, is given by ,

$$J = \sum_{n=1}^{N} \sum_{k=1}^{k} r_{nk} \|x_n - \mu_k\|^2 \tag{1}$$

The goal here is to find the values of r_{nk} and μ_k so as to minimize J. Firstly, some initial values of μ_k are chosen, then in the first satge we minimize J with respect to r_{nk}, keeping μ_k fixed. Then, it is minimized with respect to μ_k, keeping r_{nk} fixed. This two stage optimization is repeated till convergence. This minimization ith respect to r_{nk} and μ_k corresponds to the E-step and the M-step of the expectation maximization algorithm respectively. While this method involves hard assignment of the observations to various clusters, the EM algorithm involves their soft assignment to clusters i.e. each observation belongs to a cluster with a probability rather than with certainty.

Gaussian mixtures model: EM algorithm for for gaussian mixtures model is a soft version of the k-means clustering i.e here the data is assigned to different clusters with a probability rather than certainty. The Gaussian mixture distribution can be written as:

$$P(x) = \sum_{k=1}^{k} \pi_k N(x|\mu_k, \sigma_k) \tag{2}$$

In the above equation, π_k is the proportion of households belonging to each cluster k and x is the normally distributed variable with mean μ and standard deviation σ . We replace it with a binary random variable z having 1 of k representation, taking value one for element z_k and all other elements are zero. The joint distribution p(x,z) can be defined in terms of the marginal distribution p(z) and a conditional distribution p(x/z). The distribution of x is given by:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{3}$$

Supposing there are three classes(lower class,middle class and upper class in our case), the values of k=1,2,3 and the likelihood function of the household j is

given by:

$$L_{EM} = \sigma \left[\frac{\pi_1}{\sqrt{2\pi {\sigma_1}^2} e^{(x-\mu_1)^2/2{\sigma_1}^2}} \right]^{z_{1j}} \left[\frac{\pi_2}{\sqrt{2\pi {\sigma_2}^2} e^{(x-\mu_2)^2/2{\sigma_2}^2}} \right]^{z_{2j}} \left[\frac{\pi_3}{\sqrt{2\pi {\sigma_3}^2} e^{(x-\mu_3)^2/2{\sigma_3}^2}} \right]^{z_{3j}} \tag{4}$$

Since we do not know with certainty the cluster to which each household belongs, we replace z_k with the expectation of z_k for each household j given by γ_k and taking log on both sides, we get,

$$\begin{aligned} Q(.) &= \sum_{j=1}^{N} \left[\gamma_{1j} \left[log(\pi_1) - \frac{1}{2} \left(log(2\pi\sigma_1^2) + \left(\frac{x-\mu_1}{\sigma_1} \right)^2 \right) \right] \right. \\ &+ \gamma_{2j} \left[log(\pi_2) - \frac{1}{2} \left(log(2\pi\sigma_2^2) + \left(\frac{x-\mu_2}{\sigma_2} \right)^2 \right) \right] \\ &+ \left. \gamma_{3j} \left[log(\pi_3) - \frac{1}{2} \left(log(2\pi\sigma_3^2) + \left(\frac{x-\mu_3}{\sigma_3} \right)^2 \right) \right] \right] \end{aligned} \tag{5}$$

The expected values of γ at every i^{th} step is obtained using the values of π , μ and σ at the previous iteration . This is the expectation step of the EM algorithm. Then the log likelihood function obtained using these new values of γ is maximized with respect to its parameters,in order to obtain a different set of parameter values, which is turn are again used to obatin the γ values for the next iteration. This is the maximization step of the EM algorithm. These two steps are repeatedly applied in a loop till convergence is obatined. The expected value of z_{kj} for the i^{th} step of the loop given x and the values of the parameters as obtained in the previous step is given by:

$$E\left[z_{kj} \mid x, \pi^{i-1}, \mu^{i-1}, \sigma^{i-1} \right] = \frac{\pi_k^{i-1} \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}}{\sum_{k=1}^{3} \pi_k^{i-1} \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}} \tag{6}$$

The values of γ are obtained from the above equation for every household which give the probability of the household for belonging to each cluster. These

values, known as the posterior probabilities, are then inserted into the log likelihood function for the next iteration .

Binomial mixtures model: When the data is discrete and can be represented in categorical form, we can assume that it is binomially distributed and use the binomial mixtures model. The binomial mixture distribution can be written as:

$$P(x) = \sum_{k=1}^{k} \pi_k B(x \mid p, N) \tag{7}$$

In the above equation, π_k is the proportion of households belonging to each cluster k and x is the binomially distributed variable with the probability of success of any outcome as p and total number of outcomes N. We replace it with a binary random variable z having 1 of k representation, taking value one for element z_k and all other elements are zero. The joint distribution p(x,z) can be defined in terms of the marginal distribution p(z) and a conditional distribution p(x/z). The distribution of x is given by:

$$f(x|p, N) = \binom{12}{x} p^x (1-p)^{N-x} \tag{8}$$

Supposing there are three classes(lower class,middle class and upper class in our case), the values of k=1,2,3 and the likelihood function of the household j for N=12 is given by:

$$\begin{aligned} L_{EM} &= \left[log(\pi_1) + log\binom{12}{x} + log(p_1^x) + log(1-p_1)^{12-x}\right]^{z_{1j}} \\ &\times \left[log(\pi_2) + log\binom{12}{x} + log(p_2^x) + log(1-p_2)^{12-x}\right]^{z_{2j}} \\ &\times \left[log(\pi_3) + log\binom{12}{x} + log(p_3^x) + log(1-p_3)^{12-x}\right]^{z_{3j}} \end{aligned} \tag{9}$$

Since we do not know with certainty the cluster to which each household belongs, we replace z_k with the expectation of z_k for each household j given by γ_k and taking log on both sides, we get,

$$
\begin{aligned}
Q(.) &= \sum_{j=1}^{N} \Bigg[\Big[\gamma_{1j} [log(\pi_1) + log\binom{12}{x_j} + x_j log(p_1) + (12 - x_j) log(1 - p_1) \Big] \\
&+ \gamma_{2j} \Big[log(\pi_2) + log\binom{12}{x_j} + x_j log(p_2) + (12 - x_j) log(1 - p_2) \Big] \\
&+ (1 - \gamma_{1j} - \gamma_{2j}) \Big[log(1 - \pi_1 - \pi_2) + log\binom{12}{x_j} + x_j log p_3 + (12 - x_j) log(1 - p_3) \Big] \Bigg]
\end{aligned} \tag{10}
$$

The expected values of γ at every i^{th} step is obtained using the values of p and π at the previous iteration . This is the expectation step of the EM algorithm. Then the log likelihood function obtained using these new values of γ is maximized with respect to its parameters,in order to obtain a different set of parameter values, which is turn are again used to obatin the γ values for the next iteration. This is the maximization step of the EM algorithm. These two steps are repeatedly applied in a loop till convergence is obatined. The expected value of z_{kj} for the i^{th} step of the loop given x and the values of the parameters as obtained in the previous step is given by:

$$
E\Big[z_{kj} \mid x, \pi^{i-1}, p^{i-1} \Big] = \frac{\pi_k^{i-1} \binom{12}{x} p_k^x (1 - p_k)^{12-x}}{\sum_{k=1}^{3} \pi_k^{i-1} \binom{12}{x} p_k^x (1 - p_k)^{12-x}} = \gamma_{kj}^i \tag{11}
$$

The values of γ are obtained from the above equation for every household which give the probability of the household for belonging to each cluster. These values, known as the posterior probablities, are then inserted into the log likelihood function for the next iteration .

In this paper, we will be using both the binomial as well as the gaussian mixtures model for the purpose of clustering households into lower, middle and upper class for rural and urban areas respectively. It has been shown in previous studies(Maitra) that clustering data into three classes gives more consistent results compared two or more than three classes. For rural areas, three criteria

are used: total land possessed, consumption expenditure and number of durables owned by the household. In urban areas, we only use the criteria of consumption expenditure and number of durables owned for categorizing households into economic classes.The possession of 12 durable goods is considered for this purpose and the posession of any durable goods out of these 12 is assumed to be of equal weightage. Although the durables vary in terms of value and utility and the possession of each durable may be correlated to others, this assumption is made for simplification and the possession of each good is assumed to be identically and independently distributed with respect to others. The 12 durable goods used here can be divided into three categories: transport goods(scooter and car),entertainment goods(television, computer, VCD and camera) and household goods(airconditioner,washing machine,inverter,stove,refrigerator and water purifier). We assume that total land owned and monthly per capita consumption expenditure are normally distributed while the number of durables owned is binomially distributed. Once the households are clustered into these three classes on the basis of each of these criteria individually, we device a strategy to categorize the households into economic classes jointly on the basis of all three criteria for rural households and both the criteria for urban households. A rural household has to belong to a particular class according to at least two of the given three criteria in order to be assisgned that class, otherwise it will be put into a class below that. There is no ordering as such between these three criteria. Therefore, we take the sum of ranks alloted in all three clusters for each household in order to assign the final class that it belongs to. For instance, if the household belongs to lower class according to one criteria and middle class according to two, we assign that household a middle class status. Suppose, we number a lower class household as 1, a middle class household as 2 and an upper class household as 3. Since only the sum of these ranks matter, any household with the ranks summing over to 5, as in the above mentioned case will be considered middle class. For instance, a household classified as lower class according to two of the three criteria and upper class according to one criteria, will have the ranks summing over to 5 and will be considered as a middle class household. On similar lines, we cluster urban households into three classes on

the basis of consumption expenditure and durable ownership. Once the rural and urban households are assigned their respective classes, we can observe how different socio economic factors and consumption patterns behave with respect to economic class.

5 Results and discussion

In this section, we will present the results obtained through the cut off approach and the mixtures model approach and compare the findings pertaining to the year 2011-12.

5.1 Consumption expenditure approach

Using the consumption expenditure approach, it is found that 29.28% rural and 32.65% urban households fall below the poverty line as per the latest criteria of Rs.32 per capita per day and Rs.47 per capita per day for rural and urban households respectively. Also, 46.68% of the rural and 35.79% of the urban households have per capita consumption expenditure below Rs.73 and Rs.98 respectively but above the poverty line. So, there is a considerable proportion of households with per capita consumption expenditure above poverty line but are still very vulnerable to poverty. These poor and vulnerable households are jointly considered as lower class households. The middle class as defined earlier constitutes 21.33% rural and 20.42% urban households while the global middle class with an income above $ 10 but below $ 100 constitutes 2.67% rural households and 11.08% urban households. Figure 1(a) and 1(b) show the pie diagrams depicting the share of households belonging to these economic categories in rural and urban areas respectively.

Employment patterns: In rural areas, it is seen that the poor and the vulnerable(lower class) have almost similar employment pattern. A large share of these lower class households have members engaged in self employment both in agriculture and non agriculture. This indicates that many lower class households are dependent on small plots of unviable land and the produce is not enough

(a) Rural (b) Urban

Note: Authors' calculation from NSS data

Figure 1: Economic class composition

for ensuring them financial security. Many start small roadside shops or other small household business as an alternate source of income, which again may not be very profitable (Banerjee and Duflo, 2008). Among the middle class, global middle class and the rich, a considerable proportion of households are engaged as regular wage earners apart from self employment in agriculture and non agriculture. As we move from poor towards the rich economic category, we find that the proportion of households self employed in non agriculture and working as casual labourers declines. However, the proportion of households working in other occupations increases with the proportion being the highest in case of rich households. These may constitute the unproductive intermediary class of moneylenders and brokers. Self employment in agriculture is an ambiguous category due to wide disparity in the size of landholdings among farmers. Table A1 depicts the proportion of households engaged in various occupations in rural areas.

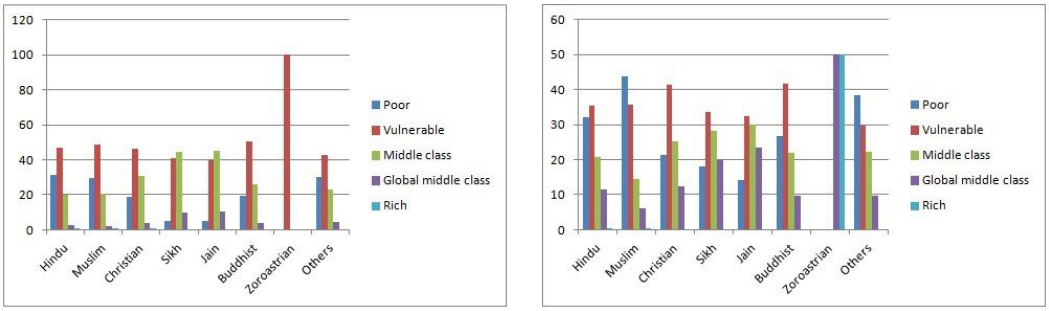
In urban areas, a larger proportion of households are engaged in self employment among the lower class and the middle class compared to the global middle class and the rich. These constitute small and micro enterprises mostly run at the household level. A considerable share of households have regular wage employment as their primary occupation across all economic categories although the share is slightly greater among the middle class and the global middle class

households. Again, the share of households involved in casual labour activities declines and the share of households engaged in other occupations increases up the economic class ladder. The results are shown in table A2.

Education: The share of households with all illiterate individuals clearly declines as we move up the scale from the poor to the rich both in rural as well as urban areas. On the other hand, the share of households with highest education level of a member as high school and above, graduate and above and post graduate and above clearly increases as we move up the economic class ladder. Only in the case of the rural rich, the share of households with these higher education levels declines as compared to the global middle class but is still higher than that for the poor, vulnerable and the middle class. These trends clearly show that households belonging to a higher economic class are more inclined towards attaining higher education levels except in case of the rural rich who are mostly traditional landed classes who may not have an incentive to acquire higher education. Also, the proportion of households with higher educational qualifications is higher across all economic categories in urban areas as compared to rural areas. These findings are shown in table A3 and table A4.

Religion: We now look at the economic class composition of different religious groups in rural and urban areas. Among rural households, it is found that around 3 out of 10 Hindu, Muslim and other households are poor, 2 out of 10 Christian and Buddhist households are poor while 1 out of 10 Sikh and Jain households are poor. About 5 in 10 Hindu, Muslim, Christian and Buddhist households are vulnerable, 4 in 10 among Sikh, Jain and other households are vulnerable while all rural Zoorashtrian households are vulnerable. Approximately, 2 out of 10 Hindu, Muslim and other households are middle class, 3 out of 10 Christian and Buddhist households are middle class and 5 out of 10 Sikh and Jain households are middle class. 1 in 10 Sikh and Jain households are global middle class while among other religious groups, less than 1 in 10 households are global middle class. A negligible proportion of households are rich across all religious categories with the highest proportion of rich among Christians.

In urban areas, approximately, 3 out 10 Hindu and Buddhist households are



(a) Rural (b) Urban

Note: Authors' calculation from NSS data

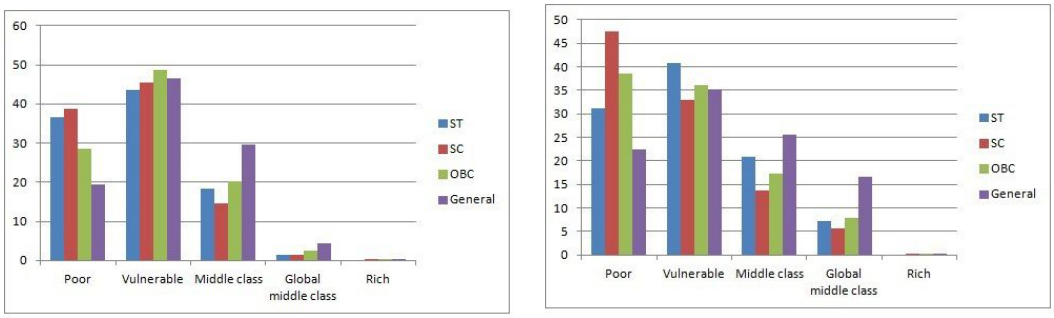
Figure 2: Economic class distribution across religious categories

poor, 2 out of 10 Christian and Sikh households are poor, 4 out of 10 Muslim and other households are poor and about 1 out of 10 Jain households are poor. About 3 out of 10 Hindu, Muslim, Sikh, Jain and other households are vulnerable and 4 out of 10 Christian and Buddhist households are vulnerable.2 out of 10 Hindu, Christian, Buddhist and other households are middle class, 3 out of 10 Sikh and Jain households are middle class while 1 out of 10 Muslim household is middle class. 1 out of 10 Hindu, Buddhist, Christian and other households are global middle class, less than 1 in 10 Muslim household is global middle class while 2 out of 10 Sikh and Jain households are global middle class and 5 in 10 Zorashtrian households are global middle class. The proportion of rich is negligible across categories except among Zorashtrians, among whom 5 in 10 households are rich.

These results are summarized in figures 2(a) and 2(b) respectively.

Caste: Next, we look at the economic class composition of different caste groups, again quoting the figures in terms of number of households for every 10 households, since it is easier to interpret. Among the scheduled tribes and scheduled castes in rural areas , about 4 in 10 households are found to be poor,3 in 10 among the OBC's are poor and 2 in 10 among general category households are poor. 4 in 10 SC and ST households are vulnerable and approximately 5 in 10 OBC and general category households are vulnerable. About 2 in 10 ST

and OBC households are middle class, 1 in 10 among SC households and 3 in 10 among general category households. Across all castes, there are less than 1 in 10 global middle class households but the general category households have the largest proportion of global middle class. Also, the proportion of rich is almost negligible across all castes except that it is slightly higher in case of general category households.



(a) Rural (b) Urban

Note: Authors' calculation from NSS data

Figure 3: Economic class distribution across social groups (castes)

In urban areas, 3 out of 10 ST households are poor, 5 out of 10 among the SC households, 4 out of 10 of the OBC households and 2 out of 10 among the general category households. 4 out of 10 ST households and approximately 3 out of 10 SC, OBC and general category households are vulnerable. 2 out of 10 ST households are middle class, less than two out of 10 SC and OBC households are middle class and more than 2 out of 10 general category households are middle class. Less than 1 out of 10 ST, SC and OBC households are global middle class and around 2 out of 10 general category households are global middle class. Again, the proportion of rich households is negligible across all categories.

These results are depicted in figure 3.

5.2 Mixtures model approach

We shall now discuss the results obtained when the mixtures model is used to identify three economic classes, namely, the lower class, middle class and upper

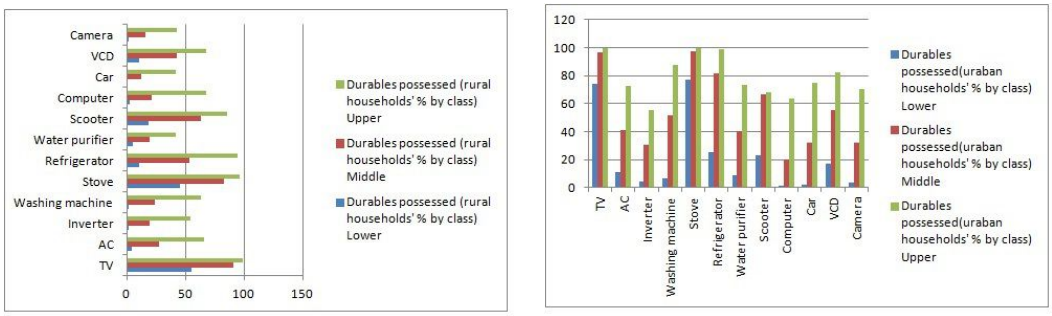
class among the rural and urban households. As discussed earlier, this model is used to cluster households on the basis of durables ownership, monthly per capita expenditure(MPCE) and possession of land in rural areas and on the basis durables ownership and MPCE in urban areas. The first step to this procedure is to cluster the household on the basis of each of these variables using the procedure of K-means clustering that gives us the initial values of the parameters in each case which are used as priors for the expectation maximisation procedure. Next, we obtain the final values of the parameters and their confidence intervals obtained after the m step of the procedure for each variable separately. Then it is found what proportion of households are clustered into each of these economic classes jointly on the basis of all these variables using the procedure describes earlier. Finally, different characteristics of these classes are studied on similar lines as has been done in case of the cut off approach.

The EM Algorithm: As explained earlier, the EM algorithm comprises of an E-step and an M-step and the final parameter values are obtained after many iterations that occur in a loop till convergence is reached. In the present case, we have assumed MPCE and land possessed to be normally distributed and hence the parameters obtained are mean, standard deviation and the proportion of households belonging to each class (say,alpha). Durable ownership is assumed to be binomially distributed and the parameters obtained in this case are probability(of ownership of any characteristic durable) and the alpha values. In rural areas, according to the MPCE criteria,60.80 % households are found to be lower class, 34.20 % are found to be middle class and around 5 % are found to be upper class. According to the durable ownership criteria, 46.10 % households are found to be lower class, 31.17 % are found to be middle class and 22.70 % are found to be upper class. According to ownership of land, 51 % of households are lower class, 33 % are middle class and 16 % are upper class. In urban areas, according to MPCE, about 46.26 % are lower class households, 43.64 % are middle class and 10.10 % are upper class and by the criteria od durables ownership, around 54 % are lower class households, 36 % are middle class and 10 % are upper class.

Once we have obtained the classification of households separately on the basis

of MPCE, durable ownership and land ownership in rural areas and on the basis of MPCE and durable ownership in urban areas, we aggregate these results on the basis of the methodology described earlier in order to assign households to different classes jointly on the basis of these criteria for rural and urban areas respectively. We find that in rural areas 80.42 % households are found to be lower class, 19.20 % households are found to be middle class and 0.38 % are found to be upper class while in urban areas, 66 % are lower class, 31.27 % are middle class and 2.80 % households are upper class. The mean MPCE in rural areas is Rs.1284.32 for lower class households, Rs.3036.64 for the middle class and Rs.8246.20 for upper class households. In urban areas, it is Rs. 1629.04 for lower class households, Rs.3947.42 for middle class households and Rs.10911.90 for upper class households. Also, 28 % of lower class ,62 % middle class and 81 % upper class households in rural areas have a member who has passed high school. In urban areas, the corresponding figures are 40.68 %, 79.50 % and 99.10 % respectively. Similarly, the percentage of households having a member who has passed graduation increases successively from lower class to upper class both in rural and urban areas. The household size increases from lower class to upper class in rural areas while it declines in urban areas indicating that richer households in urban areas have more incentives for reducing fertility rates. Expenditure on food as a share of total consumption expenditure successively declines for richer households both in rural as well as urban areas following Engel's law. These results are summarised in table A5 and A6. Also, figure 4(a) and 4(b) show the ownership pattern of individual durable goods in rural and urban areas respectively.

Employment: In rural areas, majority of the upper class households are self employed in agriculture, among middle class, the largest proportion of households have members engaged as regular wage employees and a greater proportion of lower class households are self employed in non agriculture or working as casual labourers. In urban areas, a larger proportion of lower class households are engaged in self employment as compared to the middle class and the upper class. Majority of the urban upper class households and a large proportion of middle



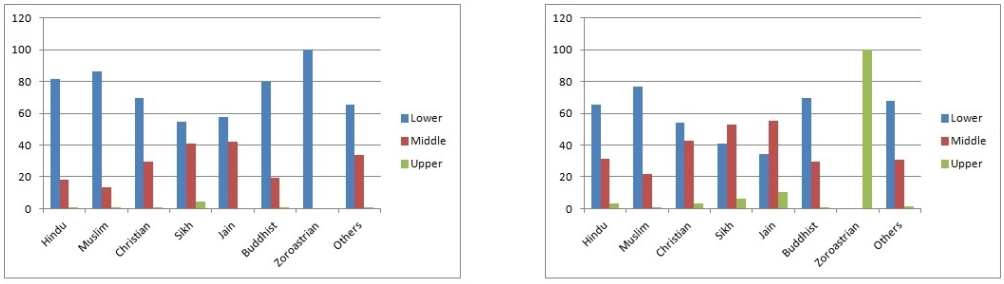
(a) Rural (b) Urban

Note: Authors' calculation from NSS data

Figure 4: Pattern of ownership of individual durable goods

class households are regular wage earners. A much greater percentage of lower class households have members working as casual labourers than middle class and upper class households. These results are shown in table A7 and A8.

Religion: In rural areas, around 8 in 10 Hindu, Muslim and Buddhist households are lower class, approximately 7 in 10 Christian and other households are lower class about 6 in 10 Jain and 5 in 10 Sikh households are lower class and all Zoroashtrian households are lower class. About 2 in 10 Hindu and Buddhist households are middle class, 1 in 10 Muslim household is middle class, 4 in 10 Sikh and Jain households are middle class and 3 in 10 Christian and other households are middle class. The proportion of upper class is negligible across religions but is slightly higher for Sikhs. In urban areas, 7 out of 10 Hindu, Buddhist and other households are lower class, 8 out of 10 Muslim households are lower class, 5 out of 10 urban households are lower class among Christian households, 4 out of 10 among Sikhs and 3 out of 10 among Jains. Also, 3 out of 10 Hindu, Buddhist and other households are middle class, 4 out of 10 Christian ,2 out of 10 Muslim and 5 out of 10 Sikh and Jain households are middle class . 1 in 10 Jain households and all Zoroashtrian households are upper class and the proportion of upper class among all other urban religious groups is less than 1 in 10. These results are shown in figure 5(a) and 5(b).



(a) Rural (b) Urban

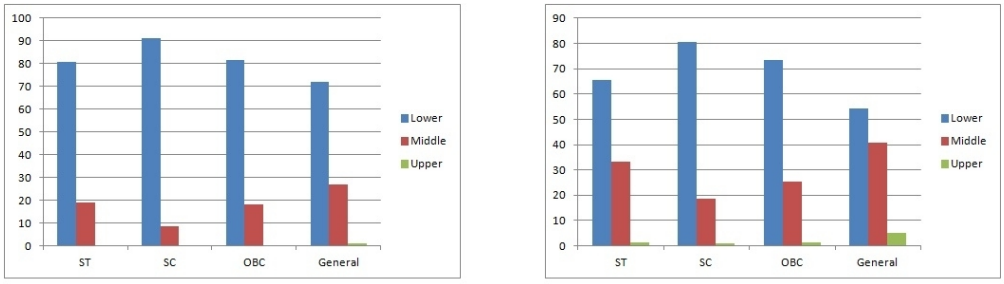
Note: Authors' calculation from NSS data

Figure 5: Economic class composition of different religious groups in mixture model

Caste: In rural areas, 8 in 10 ST and OBC households are lower class, 9 in 10 among SC households are lower class and 7 out of 10 general category households are lower class. 2 in 10 ST and OBC households are middle class, 1 in 10 among SC and 3 in 10 among general category households are middle class. Negligible proportion of households are upper class across all caste categories, the proportion being highest among the general category households. In urban areas, about 7 in 10 ST and OBC households are lower class, 8 in 10 among SC households and 5 in 10 among general category households. Also, 3 in 10 ST and OBC households are middle class, 2 in 10 SC households and 4 in 10 general category households are middle class. Again the proportion upper class households is very small across all castes and it is highest for general category households and lowest for SC households. These results are shown in figure 6(a) and 6(b).

5.3 Comparing the results of the two approaches

We have now classified the households into economic classes and identified the middle class using both the per capita consumption expenditure cut off approach as well as the mixtures model approach. While the former approach is based on certain cut offs, the latter is a probabilistic approach that assigns households to different clusters based on the probability of a household belonging to a certain



(a) Rural (b) Urban

Note: Authors' calculation from NSS data

Figure 6: Economic class composition of different social groups in mixture model

cluster.Also, the cut off approach is based on a single criteria ,that is, MPCE while the mixtures model takes into account the MPCE, durables owned and the total land possessed. We find that 76 % of the rural households are lower class(the poor and vulnerable combined) according to the cut off approach while 80% rural households belong to lower class according to the mixtures model . Also, 23 % rural households are middle class(the middle class and the global middle class combined) by the former approach and 19.20% are middle class according to the latter. 0.04 % households are rich by cut off approach and 0.38 % are rich(upper class) by the mixtures model approach in rural areas. In urban areas, the corresponding figures are 68 % and 66 % for lower class, 31.50 % and 31.27 % for middle class and 0.06 % and 2.68 % for the upper class households. These figures show that the results from both these approaches are more or less comparable.The proportion of households that belong to middle class using either of these approaches is not very divergent. However, the proportion of rich using the mixtures model approach is almost 9 times that using the cut off approach in rural areas and almost 45 times in urban areas. Although the proportion of rich is small using either of these approaches, the difference may be substantial in absolute terms.

Using both the methodologies,we find that the share of households with higher education levels increases as we move from the poorer to the richer households (except in case of the rural rich using the cut off approach). It is found that the the proportion of rural middle class and rural global middle class households

possessing graduation and higher degrees is higher than the proportion of rich households with similar qualifications using the cut off approach. Further, it is observed that the share of middle class and upper class households with high school and graduation degrees is higher when the mixtures model approach is used than the share we obtain using the cut off approach. As far as employment patterns are concerned, the cut off approach indicates that a large proportion of rural lower class households are self employed in agriculture. This share is much lower in case of the rural lower class and even middle class households classified using the mixtures model approach. It is found that 26 % rural middle class households are engaged in regular wage employment compared to 14.28 % rich rural households when the mixtures model is used. The corresponding shares using the cut off approach are 29 % and 34.62 % respectively. Using either of these approaches, it is seen that the Sikhs and Jains have a greater proportion of households belonging to middle class in both rural and urban areas. In rural areas, the economic class composition of religious groups is more or less consistent using both the approaches except that the proportion of Muslim lower class households is slightly lower using the cut off approach than that using the mixtures model approach. Among urban households, the proportion of middle class among Christian and Sikh households is higher using the mixtures model than the cut off approach. As far as the economic class composition of different castes is concerned,the proportion of middle class households is higher among the general category households using both the approaches. The proportion of rural OBC and general category households belonging to lower class is slightly higher using the mixtures model approach. On the other hand, the proportion urban general category households belonging to lower class is slightly higher using the mixtures model approach. Apart from these slight discrepancies, the socio-economic behaviour of the middle class is found to to be very similar upon comparing these two approaches.

6 Conclusion

In this paper, we have identified the middle class and other economic classes based on two approaches: the MPCE cut off approach and the mixtures model approach. The results from both these approaches show that the vast majority of households are lower class, a relatively smaller percentage are middle class and a negligible proportion of households are rich.According to the cut off approach,around 23% households are middle class in rural areas and 31.50 % in urban areas in 2011-12. The findings of the probabilistic approach show that 19.20 % households are middle class in rural areas and 31.27 % in urban areas for the same period. It is observed that a greater share of middle and upper class households have highly educated members compared to lower class households according to both these approaches. The economic class composition of different castes shows that the lower castes have a greater proportion of lower class households while the upper castes have a larger share of middle class and upper class households. This result is consistent in the findings using either of the two approaches. The proportion of lower class, middle class and upper class is very similar in both these approaches except minor differences. Regarding the robustness of methodology, it is difficult to comment since both approaches have their pros and cons. Howerver, the probabilistic approach makes the assignment of classes less rigid and is not based on any assumptions regarding the the range of the distribution within which each class must lie.The pre-determined cut offs are usually not very reliable since it is very difficult to establish a certain benchmark for each class for a particular region and time. The mixtures model used in this paper is based on a larger set of variables and uses relevant variables for rural and urban sectors respectively. The results from mixtures model are consistent and comparable to those obtained using the cut off approach. Therefore, the classification of households as middle class using mixtures model seems fairly thorough given the constraint of data availability and can be used for further scrutinizing different characteristics of households across economic categories.

References

NCAER report (2010). Ncaer annual report.

Banerjee, A. and Piketty, T. (2005). Top indian incomes, 1922–2000. *The World Bank Economic Review*, 19(1):1–20.

Banerjee, A. V. and Duflo, E. (2008). What is middle class about the middle classes around the world? *The Journal of Economic Perspectives*, 22(2):3–41A.

Beinhocker, E. D., Farrell, D., and Zainulbhai, A. S. (2007). Tracking the growth of india's middle class. *McKinsey Quarterly*, 3:50.

Birdsall, N. (2010). The (indispensable) middle class in developing countries. *Equity and growth in a globalizing world*, 157.

Birdsall, N., Graham, C., and Pettinato, S. (2000). Stuck in tunnel: Is globalization muddling the middle? *Brookings Institution Center Working Paper*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Easterly, W. (2001). The middle class consensus and economic development. *Journal of Economic Growth*, 6(4):317–335.

Fernandes, L. (2000). Restructuring the new middle class in liberalizing india. *Comparative Studies of South Asia, Africa and the Middle East*, 20(1):88–104.

Kharas, H. (2010). The emerging middle class in developing countries.

Kharas, H. (2011). The rise of the middle class. In Ghani, E., editor, *Reshaping tomorrow:Is South Asia ready for the big leap?*, chapter 3, pages 57–82. Oxford University Press, Oxford.

Lopez-Calva, L. F. and Ortiz-Juarez, E. (2014). A vulnerability approach to the definition of the middle class. *The Journal of Economic Inequality*, 12(1):23–47.

Maitra, S. (2007). Who are the indian middle class? an em approach using durables ownership.

McLachlan, G. and Jones, P. (1988). Fitting mixture models to grouped and truncated data via the em algorithm. *Biometrics*, pages 571–578.

Meyer, C. and Birdsall, N. (2012). New estimates of indias middle class. *CGD Note, Center for Global Development, Washington, DC*.

Milanovic, B. and Yitzhaki, S. (2002). Decomposing world income distribution: Does the world have a middle class? *Review of Income and Wealth*, 48(2):155–178.

Patnaik, U. (2010). trends in urban poverty under economic reforms: 1993-94 to 2004-05. *Economic and Political Weekly*, pages 42–53.

Pew Research Center (2012). The lost decade of the middle class.

Pierre, B. (1979). La distinction. critique sociale du jugement. *Paris, Ed. de Minuit*.

Rangarajan committee report (2014). Report of the expert group to review the methodology for measurement of poverty.

Ravallion, M. (2010). The developing worlds bulging (but vulnerable) middle class. *World Development*, 38(4):445–454.

Swaminathan, M. (2010). The new poverty line: a methodology deeply flawed. *Indian Journal of Human Development*, 4(1):121–125.

Tendulkar committee report (2009). Report of the expert group to review the methodology for measurement of poverty.

Thurlow, J., Resnick, D., and Ubogu, D. (2015). Matching concepts with measurement: Who belongs to africa's middle class? *Journal of International Development*, 27(5):588–608.

Wood, G. (1999). Binomial mixtures: geometric estimation of the mixing distribution. *Annals of statistics*, pages 1706–1721.

Table A1: Share of various economic classes in different rural employment categories(%)

Economic class	Rich	Global middle class	Middle class	Vulnerable	Poor
Self employed in agriculture	26.92	32.00	28.00	29.11	28.12
Self employed in non agriculture	19.23	18.35	24.75	26.67	25.63
Regular wage earner	34.62	33.00	29.00	17.69	17.94
Casual labour in agriculture	3.85	1.66	3.25	7.56	8.19
Casual labour in non agriculture	0.00	4.00	7.48	14.14	14.68
Others	15.38	11.00	7.55	4.84	5.44

Table A2: Share of various economic classes in different urban employment categories(%)

Economic class	Rich	Global middle class	Middle class	Vulnerable	Poor
Self employment	28.00	26.30	31.75	37.87	37.29
Regular wage	48.00	52.32	50.19	41.77	39.02
Casual labour	4.00	2.25	3.50	10.62	12.84
Others	20.00	19.11	14.56	9.74	10.85

Table A3: Share of households with various levels of education(% of rural households)

Education level	Poor	Vulnerable	Middle class	Global middle class	Rich
Illiterate	43.00	27.84	16.56	10.55	3.85
High school	5.50	14.00	28.00	42.00	34.50
Graduate	2.00	6.00	14.50	25.00	11.50
Post graduate	0.31	1.00	3.67	8.00	3.85

Table A4: Share of households with various levels of education(% of urban households)

Education level	Poor	Vulnerable	Middle class	Global middle class	Rich
Illiterate	29.46	12.59	4.72	2.43	0.00
High school	12.00	32.50	56.00	70.50	88.00
Graduate	5.50	17.00	36.00	51.00	68.00
Post graduate	1.00	4.00	10.68	19.60	32.00

Table A5: Key features of the rural economic classes

	Lower class	Middle class	Upper class
MPCE(mean in rupees)	1284.32	3036.64	8246.20
Percentage households	80.42	19.20	0.38
Household size	4.70	5.09	5.35
Education(high school(%))	27.87	61.78	80.80
Education(graduation(%))	10.83	34.35	50.00
Food expenditure(% of total)	56.70	46.40	34.04

Table A6: Key features of the urban economic classes

	Lower class	Middle class	Upper class
MPCE(mean in rupees)	1629.04	3947.42	10911.90
Percentage households	65.95	31.27	2.78
Household size	4.44	3.98	3.45
Education(high school(%))	40.68	79.50	99.10
Education(graduation(%))	19.32	57.08	84.16
Food expenditure(% of total)	53.10	39.73	25.52

Table A7: Employment patterns of different economic categories(Rural)

	Lower class	Middle class	Upper class
self employed in agriculture	23.56	46.36	68.75
self employed in non agriculture	27.43	18.33	14.73
Regular wage/salary earner	16.06	25.80	14.28
Casual labour in agriculture	9.90	1.22	0.00
Casual labour in non agriculture	17.60	2.66	0.00
Others	5.40	5.60	2.23

Table A8: Employment patterns of different economic categories(Urban)

	Lower class	Middle class	Upper class
self employed	38.46	35.50	29.28
Regular wage/salary earner	33.25	49.75	55.48
Casual labour	18.15	2.71	0.77
Others	10.14	12.05	14.47