Impact of poverty on women's nutritional status in India
Application of finite mixture models

Pontes, Leandro P.
ISCTE – Lisbon University Institute, UNIDE
Av. das Forças Armadas
1649-026 Lisboa, Portugal
lmbpontes@gmail.com

Padmadas, Sabu S.
University of Southampton. Centre for Global Health, Population, Poverty & Policy
Southampton, United Kingdom
S.Padmadas@soton.ac.uk

Dias, José G.
ISCTE – Lisbon University Institute, Department of Quantitative Methods
Av. das Forças Armadas
1649-026 Lisboa, Portugal
jose.dias@iscte.pt

This paper analyses the effect of household socio-economic position on the nutritional condition of Indian women, using data from the 2005-06 National Family Health Surveys. India along with other transition countries has experienced a steady economic growth since the last decade. Yet, there is little improvement in nutritional and wellbeing indicators particularly of women and children. There is clearly evidence of a nutrition transition in India, which is accompanied by an overall increase in average Body Mass Index leading to obesity, however, simultaneously with persistent high levels of under nutrition in large segments of the Indian society. The reasons behind this double burden have not been investigated systematically. On the other hand, the existing public health policies in India have not worked effectively in targeting nutritionally deprived population groups. The analysis considered Body Mass Index (BMI) as an outcome indicator of nutritional status, treated as continuous variable and its probability distribution is estimated by a Gaussian mixture model. The result is a mixture of Gaussian distributions, one for each cluster, with the mean being explained by a set of covariates of the socio economic status of the individuals.

Introduction

The recent economic growth in India has unprecedented effect on the living conditions and nutritional outcomes of its population. Within the nutritional transition framework, the pattern evolving in most transition economies is an increase in the number of overweight individuals, first among individuals belonging to higher socioeconomic status and progressively expanding to lower strata, while groups of individuals remain in a sub-nutritional condition. This dualism of the nutrition status is a new challenge for the design and implementation of public health policies. The population groups affected by under nutrition and obesity should be the focus of public health policies. Identifying the variations in nutritional outcomes among different socioeconomic groups is critical for designing effective policies and implementing targeted interventions. The most vulnerable population group is women of reproductive ages who suffer from a range of nutritional disorders related to pregnancy and reproduction as well as those resulting from the physical burden of work within and outside households and childcare. There is increasing evidence that undernutrition and overnutrition of mothers tend to have negative influence on their own health and that of children, including risks of obesity, diabetes and cardiovascular disease.

This paper analyses the effect of household socio-economic position on the nutritional condition of Indian women, using data from the third round of the National Family Health Surveys (NFHS-3) conducted during 2005-06. NFHS-3 collected a wide range of demographic and socioeconomic data along with specific health status and health care related to children and women of reproductive ages. The household schedule included questions on various dimensions of household economic conditions such as durable, assets, livestock and landholdings – which allows to estimate the current status of household wealth. In addition, the survey measured weight and height of individual women within the household which allows to estimate the individual Body Mass Index (BMI=kg/m^2). The BMI is an indicator for the nutritional status: underweight (BMI below 18.5), overweight (above 25 and below 30) and obesity (above 30). The occupation of the respondent and level of education were included as explanatory variables and indicators of socioeconomic status within the regression framework. Other additional control variables considered include age, region, religion, caste and the number of children.

Finite mixture model

The finite mixture model assumes that the sample of observations derives from a population with a finite number (S) of groups/components, G_s, $s = 1\dots, S$ of unknown proportions:

$$f(y_i; \boldsymbol{x}_i, \boldsymbol{\varphi}) = \sum_{s=1}^{S} \pi_s f_s(y_i; \boldsymbol{x}_i, \boldsymbol{\varphi}),$$

where the proportions or weights of each component, π_s, correspond to the *a priori* probability of an observation belonging to component s, with: $\sum_{s=1}^{S} \pi_s f_s(y_i ; \boldsymbol{x}_i, \boldsymbol{\varphi}) = 1$, $\pi_s > 0$, $s = 1, \dots, S$, and $f_s(y_i; \boldsymbol{x}_i, \boldsymbol{\varphi})$ is the conditional distribution within latent class s. In our case, we assume that given that the individual i belongs to component s, each observation is characterized by the Gaussian density probability function:

$$f_s(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_s, \sigma_s^2) = \frac{1}{\sqrt{2\pi\sigma_s^2}} exp\left[-\frac{1}{2}\left(\frac{y_i - \mu_{is}}{\sigma_s}\right)^2\right]$$

in which $\mu_{is} = \boldsymbol{x}_i\boldsymbol{\beta}_s$ and σ_s^2 are the mean and variance, respectively. The vector $\boldsymbol{x}_i$ contains one for the intercept and the covariates for observation i, and β_{ls} measures the impact of the explanatory variable l on the y_i in component s, β_{0s} is the constant in component s.

The estimation of the parameters $\boldsymbol{\varphi} = (\boldsymbol{\pi},\ \boldsymbol{\beta},\ \sigma^2)$, by the maximum likelihood method, results in maximizing of the likelihood function:

$$L = \prod_{i=1}^{n} f(y_i; \boldsymbol{x}_i, \boldsymbol{\varphi}) = \prod_{i=1}^{n}\sum_{s=1}^{S} \pi_s f_s(y_i; \boldsymbol{x}_i, \boldsymbol{\varphi}) = \prod_{i=1}^{n}\sum_{s=1}^{S} \pi_s \frac{1}{\sqrt{2\pi\sigma_s^2}} exp\left[-\frac{1}{2}\left(\frac{y_i - \mu_{is}}{\sigma_s}\right)^2\right].$$

An elegant way to obtain the estimates of finite mixture models is by using the EM algorithm. The EM algorithm consists in an iterative process, which can be resumed in two steps: step E (*Expectation step*) and step M (*Maximization step*). Given its properties, such as simplicity, easy implantation and numeric stability (Dias and Wedel, 2004), it is the most frequently used process for the maximization of the maximum likelihood in mixture models.

This algorithm works in an augmented space, in which a latent variable (z_{is}) is introduced, indicating whether individual i belongs to the latent component s. The variable assumes the value 1 if the individual i belongs to component s and the value 0, if not. Admitting that $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iS})$ are independent, identically distributed and follow a multinomial distribution with probability $\boldsymbol{\pi} = (\pi_1, \ldots., \pi_S)$ the log-likelihood for the complete data is:

$lnL_c = \sum_{i=1}^{n}\sum_{s=1}^{S} z_{is} \ln f_s\,(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_s, \sigma_s^2) + \sum_{i=1}^{n}\sum_{s=1}^{S} z_{is} \ln \pi_s.$

Step E in the EM algorithm consists in the calculation of $E(\ln L_c)$, the expected value of the function $\ln L_c$, in order to the distribution of the non-observed variable, z_{is}, conditioned on the observed variables y_i and $\boldsymbol{x}_i$ and on the estimate of the parameters $\boldsymbol{\varphi}$. The value of $E(\ln L_c)$ is obtained by substituting the z_{is} by its expected value:

$$E(z_{is}|y_i, \boldsymbol{x}_i, \boldsymbol{\varphi}) = P(z_{is} = 1|y_i, \boldsymbol{x}_i, \boldsymbol{\varphi}\,) = \frac{\pi_s\, f_s\,(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_s, \sigma_s^2)}{\sum_{r=1}^{S} \pi_r f_r(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_r, \sigma_r^2)},$$

which is equivalent to the *a posteriori* probability denoted by α_{is}, that is, the probability of an individual i belonging to component s conditional on the observed data.

On the other hand, the M step of the algorithm consists in maximizing the complete log-likelihood, in order to the parameters $\boldsymbol{\varphi}$, using the Newton-Raphson algorithm.

Once the algorithm has converged, parameter estimates ($\hat{\boldsymbol{\varphi}}$) characterize the mixture components; while *a posterior* probability estimates ($\hat{\alpha}_{\mathrm{is}}$) identify the best probabilistic partition of the population. The classification of each individual into the S components is based on the highest probability of an individual belonging to a certain mixture component.

The choice of the adequate number of components (S) is usually based on the utilization of criteria/statistics of the theory of information. The basic principle of these criteria is parsimony, which results in a trade-off between the complexity of the model (measured by the number of parameters) and its fit. Of the many existing criteria (measured by the maximum log-likelihood value), the BIC – Bayesian

Information Criterion (Schwarz, 1978) is the most popular in mixture modeling. It selects the number of components based on the minimization of the following criterion:

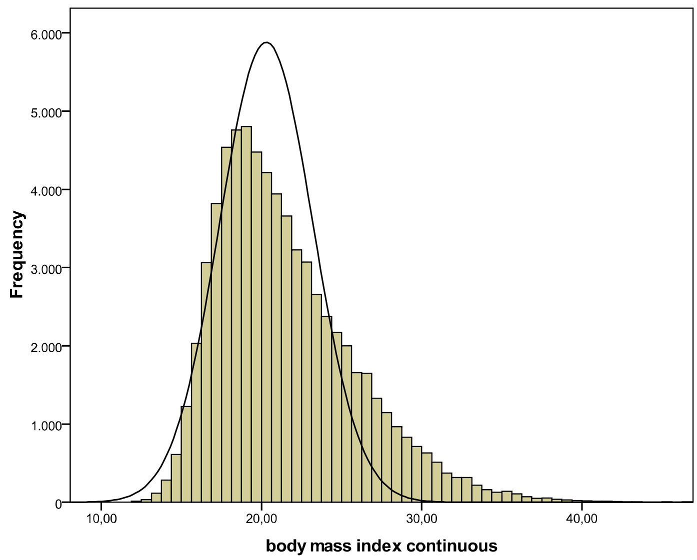
$$C_S = -2\ell_S(\widehat{\boldsymbol{\varphi}}; \boldsymbol{y}, \boldsymbol{x}) + N_S \ln(n),$$

in which $\ell_S(.)$, N_S and n correspond to the log-likelihood function, the number of parameter of the model, and the sample size, respectively. Lower values of BIC means more parsimonious models.

Results

This section presents the main results of this research. First, we analysis the shape of the BMI empirical distribution. As can be seen in Figure 1, it is strongly skewed that invalidates the use of Gaussian (symmetric) models.

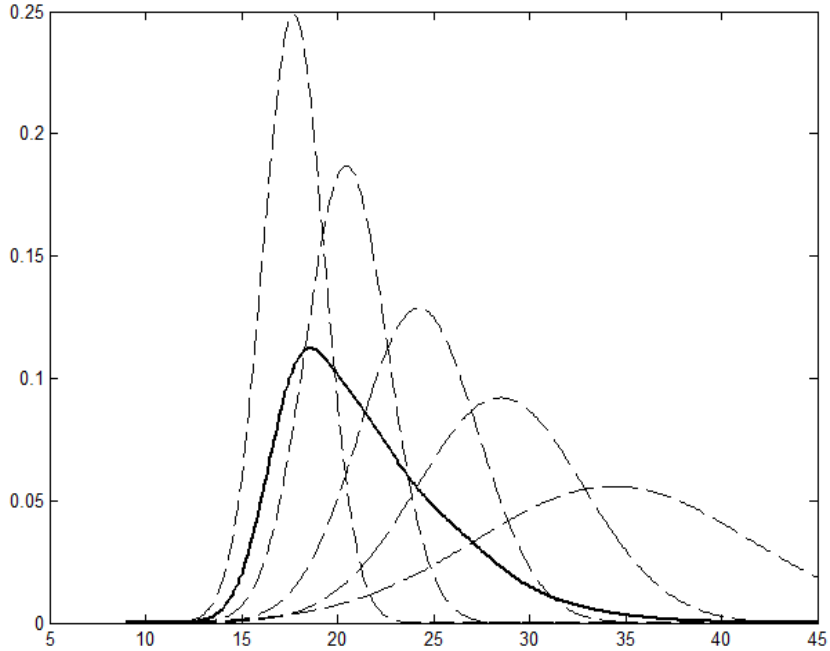
Figure 1. Empirical distribution of the BMI



variable

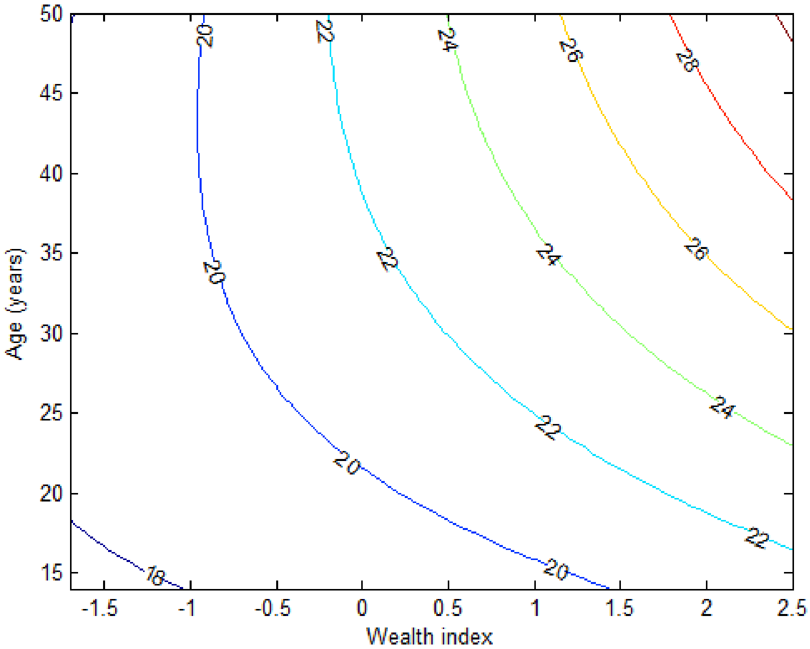
As mixture distributions can approximate any distribution, we start our modeling without covariates (only means or intercepts β_{0s} in the linear component of the model). The best model (minimum BIC) is reached with five components ($S = 5$), with LL = -191884, BIC = 383923.3, and 14 parameters (4 proportions, 5 means and 5 variances). Figure 2 depicts the estimated Gaussian components (dashed lines) and the mixture distribution that approximates the empirical distribution in Figure 1. We observe that different components retrieve different parts of the tail of the distribution. Moreover, one observes that components with increased mean tend to have larger variances. Thus, this first mixture shows that this approach can retrieve the density distribution of the dependent variable.

Figure 2. BMI distribution as a mixture of Gaussian distributions



A second model – a mixture regression model – relates the dependent (BMI) with the wealth index and age. The best BIC solution contains 6 components (LL = 180612; Npar = 47). Figure 2 sh ows the surface of the relation between the estimated expected BMI as function of wealth index and age: $E(\widehat{y_\iota|\boldsymbol{x}_\iota}, \widehat{\boldsymbol{\varphi}}) = \sum_{s=1}^{S} \widehat{\pi_s}\widehat{\mu_{\iota s}}$. In this mixture regression we include the quadratic effects and the interaction between wealth index and age in the linear component of the model. One observes that obesity increases strongly with wealth index and less pronounced with age. Undernourishment is particular expected for lower wealth index values and young ages.

Figure 3. BMI expected value from a mixture of Gaussian regressions



Our main goal is to understand the impact of the socio economic dimension on the BMI score. We performed a third mixture analysis controlling the variables Wealth Index (household), Age of respondent, Region, Religion, Caste, Number of children, Educational level of the respondent, and Occupation of respondent. The mixture of Gaussian regression that best fits the data (BIC) contains four latent classe s (LL = -181138.96; Npar = 127). Table 1 reports the estimates of the model.

Table 1. Mixture regression estimates

	Latent classes				Wald	p-value
	1	2	3	4		
Intercept	17.908 ***	18.837 ***	17.743 ***	21.705 ***	83284.3	0.000
Wealth Index (household)						
Poorest	-	-	-	-	3312.4	0.000
Poorer	0.037	0.736 ***	-0.108	0.665		
Middle	0.558 ***	1.631 ***	-0.118	2.378 ***		
Richer	1.536 ***	3.245 ***	0.030	4.336 ***		
Richest	3.452 ***	5.489 ***	0.951 ***	7.338 ***		
Age of respondent						
15-24	-	-	-	-	1322.1	0.000
25-34	0.393 ***	1.487 ***	-0.133 *	2.140 ***		
35-49	1.207 ***	2.868 ***	-0.154 *	3.835 ***		
Region						
North	-	-	-	-	800.3	0.000
Central	-0.102	-0.055	-0.151 *	-0.749 **		
East	-0.211 **	-0.493 ***	-0.205 **	-1.816 ***		
Northeast	0.258 **	-0.162	0.441 ***	-2.276 ***		
West	-0.435 ***	-0.358 ***	-0.764 ***	-0.291		
South	0.323 ***	0.845 ***	-0.245 ***	1.133 ***		
Religion						
Hindu	-	-	-	-	446.7	0.000
Muslim	0.264 ***	0.828 ***	-0.179 **	1.302 ***		
Christian	0.728 ***	0.466 **	0.594 ***	0.479		
Other	0.988 ***	1.226 ***	0.467 ***	2.459 ***		
Caste						
General	-	-	-	-	154.9	0.000
Sched caste	-0.266 ***	-0.315 ***	-0.288 ***	-0.403		
Sched tribe	-0.228 *	-0.924 ***	0.168	-1.314 ***		
OBC	-0.067	-0.322 ***	-0.089	-0.598 **		
Other/dk	-0.171	-0.228	0.144	-1.578 ***		
Number of children						
4+	-	-	-	-	44.3	0.000
3	-0.009	0.113	0.203 ***	-0.002		
2	0.135 *	0.129	0.116	0.166		
None or one	-0.110	0.063	-0.021	0.281		
Level of education of respondent						
No education	-	-	-	-	159.9	0.000
Primary	0.248 ***	0.243 **	0.062	0.417		
Secondary	0.381 ***	0.467 ***	-0.014	0.323		
Higher	0.988 ***	0.572 ***	0.386 ***	0.810 *		
Occupation of respondent						
Not working	-	-	-	-	366.5	0.000
Agricultural	-0.276 ***	-0.868 ***	0.275 ***	-1.944 ***		
Skilled,unskilled manual	-0.262 ***	-0.562 ***	-0.027	-1.146 ***		
Services	0.200	0.290	0.055	0.092		
Clerical, sales	0.345 **	0.384 *	0.650 ***	0.187		
Professional, tech, managerial	0.328 *	0.075	0.597 ***	-1.350 *		
Variance	3.412	6.404	2.663	18.292		
Latent class sizes	0.368	0.334	0.219	0.080		

The richest category of wealth index has the greatest effect on the BMI for all the classes, especially on the fourth class with the greatest BMI average. For this class, richest individuals have a BMI 7 points higher than the individual in the reference category. The BMI grows as the richness of the individuals increases for all classes. The level of education has a similar pattern of effects on BMI, however with low coefficients. Higher levels of educations have the greatest effect on BMI.

The results also suggest that living in North or South regions has a positive effect on BMI. In general, the coefficients for the other four regions have negative signals.

Older women have higher BMI in all classes except the third class. For the third class as individuals grow older their BMI decreases. This class presents the lowest average BMI.

The lifestyle has effects on BMI because it affects the balance between the dietary energy intake and the labor demands of energy. Manual labor requires more energy expenditure and so the coefficients of the individuals employed in agriculture and manual labor exhibit negative significant coefficients in almost all the classes.

Table 2. Latent class profiling

	Latent classes				**Aggregate**
	1	**2**	**3**	**4**	
Class size	0.368	0.334	0.219	0.080	
BMI (mean)	20.364	23.620	17.928	27.165	21.538
Wealth Index (household)					
Poorest	0.127	0.116	0.136	0.106	0.124
Poorer	0.151	0.144	0.160	0.137	0.150
Middle	0.191	0.190	0.196	0.188	0.192
Richer	0.232	0.239	0.229	0.244	0.235
Richest	0.299	0.311	0.278	0.326	0.301
Age of respondent					
15-24	0.149	0.134	0.173	0.121	0.147
25-34	0.400	0.392	0.413	0.384	0.399
35-49	0.451	0.474	0.414	0.496	0.454
Region					
North	0.193	0.194	0.194	0.197	0.194
Central	0.189	0.185	0.199	0.182	0.189
East	0.158	0.157	0.158	0.152	0.157
Northeast	0.135	0.128	0.140	0.124	0.133
West	0.131	0.132	0.130	0.137	0.131
South	0.193	0.204	0.180	0.209	0.195
Religion					
Hindu	0.763	0.760	0.765	0.758	0.762
Muslim	0.124	0.130	0.120	0.135	0.126
Christian	0.057	0.055	0.061	0.053	0.057
Other	0.055	0.055	0.054	0.054	0.055
Caste					
General	0.349	0.357	0.334	0.370	0.350
Sched caste	0.168	0.167	0.171	0.166	0.168
Sched tribe	0.110	0.101	0.121	0.090	0.108
OBC	0.329	0.329	0.334	0.329	0.330
Other/dk	0.044	0.046	0.041	0.046	0.044
Place of residence					
Countryside	0.582	0.542	0.612	0.490	0.568
Town	0.152	0.157	0.144	0.160	0.153
Small city	0.072	0.079	0.069	0.083	0.075
Capital, Large city	0.194	0.223	0.175	0.267	0.205

The first and second latent classes represent about 70% of the population and they both have an average BMI inside the cut-offs usually adopted for a balanced nutritional status (18<BMI<25). The third cluster represents about 22% of individuals with the lowest average BMI of 17.9 kg/m^2. Individuals residing in countryside have the greatest probability of belonging to this cluster. Also, there is a slight predominance of individuals with age between 15 and 24 years old if we compare with the other clusters.

The fourth cluster represents just 8% of total individuals. With the highest average BMI of about 27 kg/m^2, the latent class includes those with overweight. The individuals of this cluster are older than the population aggregate and live more in large cities than the individuals of the other classes.

Conclusion

Our findings shed light on the explanation of overweight and underweight based on socio economic status. Wealth, age and education and absence of a working activity have negative effects on health through the increase of the BMI. This especially true for individuals of the fourth class which present already higher average BMI than the other classes. Living in south and north regions is an additional risk factor for obesity.

The class with lowest average BMI presents a different pattern of effects. Age, belong to a scheduled cast and not working are factors that affect negatively the average BMI of the class, and represent a deterioration of health conditions.

The consequences of these findings confirm previous results which suggest an increase of BMI for the richest individuals while the individuals with low income remain with a low average BMI. It was possible to identify regions of higher risk of obesity and under nutrition. Lifestyles also have effect on BMI and sedentary life is a risk factor. Accordingly, these results suggest that not working is a risk factor for both individuals with lower BMI and also for those with higher BMI.

REFERENCES

Dias, J.G. and M. Wedel (2004), On EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods, *Statistics and Computing*, 14(4), 323-332.

Griffiths, P and M Bentley, 2005, Women of higher socio-economic status are more likely to be overweight in Karnataka, India, *European Journal of Clinical Nutrition*, 59, 1217–1220.

Padmadas, S.S., J.G. Dias and F.J. Willekens (2006), Disentangling women's responses on complex dietary intake patterns from an Indian cross-sectional survey: A latent class analysis, *Public Health Nutrition*, 9(2), 204–211.

Schwarz, G. (1978), Estimating the dimension of a model, *Annals of Statistics*, 6(2), 461-464.

Subramanian, S.V, J.M. Perkins and K.T. Khan (2009), Do burdens of underweight and overweight coexist among lower socioeconomic groups in India?, *American Journal of Clinical Nutrition*, 90, 369–76.

Subramanian, S.V. and G.D. Smith (2006), Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India1, *American Journal of Clinical Nutrition*; 84, 633–40.