A Behaviour-based Approach to the Estimation of Poverty in India*

Ingvild Almås[†] Anders Kjelsrud[‡] Rohini Somanathan[§]

Abstract

A sixth of the world's population and a large fraction of its poor live in India. Indian poverty estimates are crucial inputs in understanding world poverty, yet there is much disagreement about the numbers and the legitimacy of methods used to derive them. In this paper we propose and justify an alternative approach to identifying the poor, using the proportion of their incomes spent on food. Our estimates have weaker data requirements than official methods and compare favorably on several validation tests. Most notably, households around our state poverty lines obtain their calories from similar sources whereas this is not true of official poverty lines. We also find rates of self-reported hunger are higher in states we classify as poor (JEL: D1, E31, F01, I32).

[*] *Acknowledgements*: We thank Orazio Attanasio, Ragnhild Balsvik, Arne Bigsten, Richard Blundell, Ian Crawford, Gernot Doppelhofer, Abhiroop Mukhopadhyay, Ragnar Nymoen and Fabien Postel-Vinay for useful comments. The paper is part of the research activities at the Centre for the study of Equality, Social Organization, and Performance (ESOP) at the Department of Economics at the University of Oslo. ESOP is supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 179552.

[†] Norwegian School of Economics and University of Oslo, email: `ingvild.almas@nhh.no`

[‡] Statistics Norway, Research Department, email: `gkj@ssb.no`

[§] Dehli School of Economics, email: `rohini@econdse.org`

1 Introduction

Almost a sixth of the world's population and a large fraction of its poor live in India. Indian poverty estimates are therefore crucial inputs in understanding world poverty trends. Yet there is much disagreement about the numbers and the legitimacy of methods used to derive them. Since the nineties, separate official poverty lines have been published for urban and rural regions of each of the Indian states to reflect spatial variation in the cost of living. These estimates have been the source of considerable controversy and, over the past decade, two independent commissions have suggested new methods for estimating regional prices based on micro price data from consumption surveys. The first revision, proposed in 2009, resulted in a 50 per cent increase in rural head counts for 2004-5. The second revision, published last year, has resulted in overall poverty rates for 2011-12 that are 35 per cent higher than estimates based on the 2009 methodology. The debate on poverty measurement in India is especially charged with the political rhetoric of poverty eradication accompanying the expanding and fluctuating numbers of poor families. Also, with many government programs now targeted only to families that are officially classified as poor, correctly identifying them has assumed new importance.

Price series to adjust for cost-of-living differences are at the core of any comparisons of real income or welfare across individuals and over time, so also for poverty calculations. Several studies have suggested alternative methods of arriving at reasonable price series that could be used to generate consistent poverty estimates (Deaton, 2010; Deaton and Dupriez, 2011; Diewert, 1978; Neary, 2004; Hill, 2004). In this paper we propose and justify an alternative approach to estimating poverty that circumvents direct micro price measurement and aggregation. Since Ernst Engel's work (Engel, 1857, 1895), the empirical regularity of a negative relationship between the budget share for food and real income has been well established. We identify regional differences in the cost-of-living in India by estimating Engel curves for food. We assume that households with the same demographic and occupational characteristics spend the same proportion of their income on food. We then use data from the National Sample Surveys (NSS) and attribute systematic differences in *nominal* expenditures of households with the same food share to different relative price levels across states. We do this separately for the rural and urban samples of the NSS and then use our price estimates to derive rural and urban poverty lines and head counts for each of the Indian states in 2004-05 and 2009-10.

Our paper has two main objectives: first, to obtain a set of price and poverty estimates using the Engel approach and second, to examine whether the Engel method we use does a better job of identifying the poor than the current official methodology. If it does, then a comparison of official estimates with ours is meaningful and can reveal biases in official accounts of poverty patterns and trends in India. Our main strategy for checking the validity of our estimates is to compare the consumption behavior of households within a narrow band of our poverty lines with those in a similar band around official lines. There is evidence that the poor often get their calories from relatively cheap sources while the less poor substitute towards more expensive calories with favorable attributes such as taste or status (Jensen and Miller, 2010). We find that households clustered around our estimated lines get large and similar shares of their total calories from cheap sources such as cereals and small shares from expensive sources such as fats and oils. In contrast, for households around the official poverty lines in different states, spending continues to be related to the nominal expenditures, suggesting that these lines do not properly account for cost of living differences. We also find higher correlations between rural and urban prices than official price estimates and higher rates of self-reported hunger in the states that we classify as the poorest.

The Engel approach estimates only relative price levels.[1] To generate poverty lines that can be compared with the official ones, we normalize our estimates so as to generate the official aggregate poverty line for 2004-05. In other words, for that year, by construction, our state poverty lines weighted by the populations of the various states equal the official lines similarly weighted. Any differences between our results and official estimates therefore mainly appear in different patterns of spatial poverty in the two years, 2004-05 and 2009-10, and in changes in poverty over the five-year period. We present three main findings. There is a higher dispersion in poverty across Indian states in both years. Second, the rural poverty rates in the eastern states of Assam, Bihar, Odisha and West Bengal are consistently higher than those implied by official figures. Finally, the decrease in overall poverty over our five-year period is more modest than suggested by official statistics.

Our paper is related both to the literature on poverty measurement in India and to studies that use estimated Engel curves to correct for biases in the measurement of prices over time. Hamilton (2001) pioneered this strand of research through his study on consumer price indices in the United States and it has since been applied to several countries over varying time periods.[2] More recently, it has also been used to estimate biases in spatial price variations (Almås, 2012).[3] Studies of growth require the identification of prices over time while studies of inequality are based on spatial price variation. In order to study poverty however, the identification of *both* spatial and temporal indices are necessary and we provide a framework to do so.

[1] Our main estimates use a restricted sample of households with similar demographic composition to ensure that our results are not influenced by differences in fertility rates and family size. They appear robust to alternative empirical specifications of the Engel relationship and changes in the composition of our sample.

[2] For example, Almås and Johnsen (2013); Barrett and Brzozowski (2010); Beatty and Larsen (2005); Carvalho Filho and Chamon (2006); Chung *et al.* (2010); Costa (2001); Gibson *et al.* (2008); Larsen (2007); Nakamura and Liu (2013); Olivia and Gibson (2012) have all applied this method in different contexts.

[3] The Engel methodology has been discussed in several papers, see e.g., Deaton and Dupriez (2011) and Beatty and Crossley (2012), and validations have been called for, see e.g., Ravallion (2015).

The deficiencies in official approaches to price and poverty measurement in India have been extensively discussed in a series of papers by Angus Deaton and co-authors, who have provided alternative poverty estimates based on unit-values from the NSS (Deaton and Tarozzi, 2000; Deaton, 2008, 2010). These studies were influential in bringing about changes in the official methodology which switched from using more aggregate price data to unit values obtained from the NSS micro data. Mishra and Ray (2011), Chattopadhyay (2010) and Coondoo *et al.* (2011) also provide alternative approaches for estimating prices. We see our work as complementary to this body of research. The Engel method is attractive due to its low data requirements and clear theoretical foundations. Based on our analysis of its validity in the Indian case, we believe it can be profitably used as a benchmark with which to compare alternative official and other methodologies.

The rest of this paper is organized as follows. Section 2, sketches a chronology of poverty measurement in India. In Section 3, we describe our empirical methodology and link it to the NSS data we used. In Section 4, we present our estimates of prices and corresponding poverty lines. We also examine the behavior of households around these lines in terms of the composition of calories consumed and present other validation exercises of the poverty estimates. In Section 5, we report results from a range of specification checks. Concluding remarks are provided in Section 6.

2 Indian poverty measures: a chronology

Poverty lines during the colonial period and in the decades immediately following independence were based on arbitrary and varying notions of adequacy (Srinivasan, 2007). In 1979, subsistence needs were systematically linked to nutritional needs and household spending patterns. Calorie norms of 2400 per capita per day for rural India and 2100 for urban India were adopted and the expenditure equivalents of these norms were identified through the empirical distribution of consumer expenditure from the NSS survey of 1973–74. These became the new poverty lines for rural and urban India. Although derived from household expenditure data, they were stated in terms of monthly per capita expenditures and this continues to be the current practice (Government of India, 1979).[4] Implicitly, subsistence was defined as the bundle consumed by households at these calorie levels.

Until the 1990s, no attempt was made to capture differences in prices or spending patterns across states. Poverty estimates were revised with each quinquennial NSS survey and price deflators were used to adjust for price changes over time.[5] In 1993, an expert group set up by the Planning

[4]The 1979 poverty lines were 49 and 57 rupees in rural and urban areas respectively.

[5]The choice of deflators changed over the rounds. For details, see Government of India (1993), p. 13.

Commission recommended state-specific poverty lines based on regional prices which captured the cost-of-living for poor households (Government of India, 1993). For each state, the new price deflators were the consumer price index for agricultural labourers (CPIAL) for the rural population and the consumer price index for industrial workers (CPIIW) for its urban counterpart. The updating of poverty lines was done purely on the basis of these cost estimates.

Over the years, this method lost credibility. The price data was argued to be flawed and successive poverty lines failed to preserve the original calorie norms (Deaton, 2003, 2008; Deaton and Tarozzi, 2000). Another expert committee was formed in late 2005 led by Suresh Tendulkar and new poverty lines were published in its report in 2009. The report was officially adopted by the Planning Commission in 2011 (Government of India, 2009, 2011). The Tendulkar Committee did not relate poverty lines to calories. However, for the sake of continuity, it anchored the all-India urban headcount for 2004-05 to 25.7 per cent, the official estimate under the old procedure. Using this normalisation, it then arrived at rural and urban poverty lines for each state using elaborate methods for estimating regional price variations based on the aggregation of 23 price indices for different categories of expenditure (Government of India, 2009).

The Tendulkar methodology obtains price estimates using unit values computed from the same NSS data that are used to estimate household expenditure. Although unit values may differ from prices because they do not adjust for differences in quality, it has been argued that these biases are quite small (Deaton, 1988). A more serious objection is that it is only possible to construct unit values for items for which survey data can provide meaningful quantities. This includes most food and fuel, and some clothing, but on average about 30 percent of total expenditure is excluded (a larger fraction of expenditure is typically excluded for the rich and a lower fraction for the poor, as preferences are non-homothetic). For categories such as education, health care and other services, price information was obtained from a variety of sources, including specialized surveys on health by the NSS.[6] This makes the new procedures somewhat ad hoc and difficult to replicate in the future (see e.g., Subramanian (2011)).

The methodology proposed by the Tendulkar Committee resulted in rural poverty head counts that were 50 per cent higher than previous estimates for 2004-05. In 2012, it was used by the Planning Commission to compute poverty estimates based on the NSS consumption survey of 2009-10. The methods used and the resulting estimates continued to be controversial and yet another expert group, the Rangarajan Committee, was formed in June 2012 to evaluate them. This committee's

[6]The cost of school attendance is derived from the NSS Employment and Unemployment survey; health care costs are calculated from the NSS Morbidity and Health Care survey; and prices for the remainder of households' consumption bundles (including entertainment, services and durables) are derived from the price data underlying the CPIAL and CPIIW.

report, released in 2014, re-introduced nutritional norms into calculating subsistence and arrived at an overall poverty rate of 38.2 per cent for 2009-2010; 28 per cent higher than the prevailing official rate.

In our comparison of official estimates with those resulting from the Engel method we restrict ourselves to the numbers generated by the Tendulkar Commitee report since these continue to be the official estimates. The divergent numbers by different committees do however underscore the value of an independent and parsimonious methodology to assess the poverty numbers. We present such a methodology in the next section.

3 Methods and Data

We begin by estimating the following demand system:[7]

$$m_{hst} = a + b(\ln y_{hst} - \ln P_{st}) + \theta X_{hst} + \varepsilon_{hst}. \tag{1}$$

The budget share for food is denoted by m_{hst}, y_{hst} is the nominal household expenditure level, and X_{hst} is a vector of household-specific control variables, such as demographics, religion and occupation, for household h in state s at time t. P_{st} is the composite price of consumption in state s at time t.[8]

The only unknown variable in this regression is the overall state price level P_{st}. This is also the only variable measured at the *state/year* level. Hence, it can be identified through state- and time-specific dummy variables:

$$m_{hst} = a + b \ln y_{hst} + \theta X_{hst} + \sum_{s=2}^{N} d_{s1} D_{s1} + \sum_{s=1}^{N} d_{s2} D_{s2} + \varepsilon_{hst}. \tag{2}$$

D_{st} is the state level dummy variable for state s in period t, and N is the total number of states. State 1 in period 1 is taken as the base and, hence, D_{11} is not included in the estimation. The state dummy coefficient, d_{st}, is a function of the overall state price level, P_{st}, and the coefficient

[7]This is a restricted version of the Almost Ideal Demand system (Deaton and Muellbauer, 1980a), restricted in that we assume that the budget share for food is not influenced by relative prices. This is discussed later in this section and relaxed in Section 5.

[8]If we wanted to derive household specific poverty lines, we would also include the household specific control variables in our cost-of-living estimates (see Blundell *et al.*, 1998; Pendakur, 2002; Pollak and Wales, 1981). See Dickens *et al.* (1993) for a discussion of properties of the demand system we use. Further, if we focused on other real income levels than the poverty line threshold, we would also include total expenditure in the cost-of-living estimates as preferences are non-homothetic (see Almås and Sørensen, 2012; Pendakur, 2002).

for the logarithm of household expenditures, b:

$$d_{st} = -b \ln P_{st}. \tag{3}$$

From Equation (3), it follows that the overall price level is given by:

$$P_{st} = e^{-\frac{d_{st}}{b}} \ . \tag{4}$$

This price level is measured relative to the base state in the base time period.[9]

This method is a variation on previously approaches that use Engel curves for correcting for biases in price series. Instead of using shifts in Engel curves to identify biases in existing prices as has previously been done, we directly estimate price levels through the systematic variation in Engel curves for similar households over the two years and for each Indian state (urban and rural part, respectively). The identified prices are then used to calculate real income and poverty head counts. This allows us to identify poverty trends for each of our spatial units.

This approach is attractive because it allows us to identify price variation without relying on price data. Our main specification assumes that the budget share for food is not influenced by relative prices. We relax this assumption in Section 5 as one of our many specification checks. We do this by including a measure of relative prices constructed from unit values and find almost identical results to those from Equation 1.[10] The above demand system has been shown to be consistent with utility maximization and allows for non-homothetic tastes as well as substitution in consumption (Deaton and Muellbauer, 1980b). Our robustness analysis in Section 5 contains a more general discussion of alternative functional forms and shows that a quadratic demand system generates similar results.

Although any item of consumption could work as an indicator good, food has several advantages over other potential candidates. Its income elasticity differs substantially from unity so its budget share is sensitive to the level of household real income and therefore to the price deflator for nominal income. Also, because of its perishability, expenditures in one period cannot provide a flow of consumption in another period. Finally, studies of different countries, and over different time periods, suggest that the Engel curve for food is log-linear and stable (Banks *et al.*, 1997; Beatty and Larsen, 2005; Blundell *et al.*, 1998; Leser, 1963; Working, 1943; Yatchew, 2003).

[9] This is a normalization. All results are invariant to the choice of base state and period.

[10] The evidence on the effect of relative prices on food shares in mixed. Related studies tend to find insignificant or small effects. For the United States, Hamilton reports an insignificant positive coefficient of 0.037, whereas Costa (2001) reports a significant positive coefficient of 0.006 for the period 1919-1935 and an insignificant negative coefficient of -0.008 for the period 1960-1994. Almås reports a positive and significant coefficient equal to 0.047 in her cross-country study.

The data used to estimate the above system comes from two recent rounds of the NSS conducted in 2004–05 (the 61st round) and 2009–10 (the 66th round). Our sample consists of the 30 states and union territories used in the construction of the official poverty lines.[11] Summary statistics, covering 222,558 households, are shown in Table A.1 in the appendix.[12]

As control variables we use data on household demographics, occupation, religion, land ownership, number of free meals and the age of the household head, all taken from the same NSS consumer expenditure survey. To avoid potential biases arising from variations in family composition, we restrict ourselves to households consisting of two children and two adults for our main results. This is the most frequently observed family composition in the NSS dataset but the restriction reduces our sample size by almost 90 per cent. As a robustness check, we also estimate our model using the full sample and including controls for the numbers of children and adults. All our main findings are robust to this change of sample. Because of the different occupational categories in the urban and rural sample and also because of potential unobservable differences across the sectors, we estimate all our models separately for the urban and rural samples in the NSS. We use the full sample of 30 states and union territories in all our models, but to be parsimonious, we list state-wise results only for the 17 largest states labelled as *major states* by the NSS. These cover more than 90 per cent of the Indian population.[13]

4 Results

Table 1 reports estimates from the demand model given in Equation 2. As expected from Engel's Law, the logarithm of total monthly expenditure has a negative effect on the budget share for food. The coefficients imply expenditure elasticities of +0.77 and +0.70 in rural and urban sectors respectively, which are similar to those found in previous studies (Almås, 2012; Beatty and Larsen, 2005; Carvalho Filho and Chamon, 2006; Costa, 2001).[14]

[11] We exclude the union territories of Andaman and Nicobar Islands, Chandigarh, Daman and Diu, Dadar and Nagar Haveli and Lakshadweep, which together constitute barely one per cent of the NSS sample.

[12] Consumption expenditures are recorded based on a 30-day recall period for most consumption goods and on a 365-day recall period for some items such as durable goods, education, medical expenses. The 66th NSS round is published as two separate surveys, each with different recall periods. To obtain a comparable sample for the two time periods, and for comparability with the official poverty counts, we use the "type 1" survey version. The NSS values items received in-kind at their average local retail price, while home production is evaluated at market prices net of transport costs.

[13] According to the Indian Census 2011 they accounted for 94 per cent of the population in 2011. We focus on these states for brevity and because estimates for the other 13 states are much less reliable due to small samples. For example, in rural Delhi the sample contains only 59 households.

[14] The expenditure elasticities are calculated as $1 + \frac{b}{m}$ where m is the mean food share in the sample. Note that we express the food share in percentages, so that e.g., the expenditure elasticity in rural sector is calculated as $1 + \frac{-12.628}{54.904} = 0.77$, 54.904 being the mean percent of the budget spent on food in rural sector.

Table 1: Demand system estimates

Dep. var.: Budget share for food (%)	**Rural**	**Urban**
	(1)	(2)
Log of household expenditure	-12.628	-13.759
	(0.227)	(0.191)
Observations	14258	9113
R^2	0.379	0.522

Note: Robust standard errors, clustered at the NSS first stage sample unit, are in parentheses. Additional controls are the age of the household head, the proportion of females in the household, the number of free meals taken outside the home and dummy variables for the occupation, religion and the cultivated land categories listed in Table A.1.

We use the coefficients for log expenditures and the state-year dummies to calculate prices based on Equation 4. To obtain comparable price measures for the official methodology, we divide the poverty lines by the all-India poverty line for urban and rural sectors for each time period. These spatial price indices are displayed in columns (1)-(8) of Table A.2 in the appendix. To easily compare the spatial variation in prices generated by our methods and those followed by the Indian Planning Commission, we re-weight prices for each method and year so that their population-weighted all-India average equals 100. As seen from the coefficient of variation, there is more price variation in rural as compared to urban areas under both methods and the Engel prices imply more dispersion than official measures in both urban and rural sector. Columns (9)-(12) display the state prices for 2009–10 relatively to the all-India levels in 2004–05. This allows us to investigate the intertemporal changes in prices. The Engel estimates suggest a cost-of-living increase of about 60 per cent for the five-year period or an average annual increase of approximately 10 per cent. By comparison, the implicit Planning Commission price measures indicate an overall increase of 50 per cent, corresponding to an average annual increase of approximately 9 per cent.[15]

Given these price indices, it is straightforward to compute updated poverty lines and headcounts. Since our price measures are identified only up to a normalization, we anchor our set of prices to the all-India poverty lines for 2004–05.[16] We then derive state poverty lines for both time periods by adjusting the all-India lines for 2004–05 for the estimated state prices. This procedure implies that our estimated all-India headcount ratios for 2004–05 differ from the official ones only because of different spatial prices while the headcounts for 2009–10 deviate on both spatial and intertemporal dimensions.

Table 2 presents headcounts based on the Engel analysis together with those from current and

[15] The Engel estimates indicate relatively higher cost-of-living increases for some western and south-western states, such as Karnataka, Maharashtra and Rajasthan. To draw a parallel to the previous literature, we can also compare our price increase with that reported by the official CPI. Table A.6 reports the price growth reported by CPI and reveals that we find a larger price for both urban and rural areas and hence that the CPI is, according to the Engel method, biased downwards.

[16] This normalization is attractive because it allows us to compare our measures to official ones.

previous official methods. The salient differences are as follows: First, we find more geographical variation in poverty than either of the official measures. This is true for both rural and urban sectors, and both time periods. Second, there are consistently higher concentrations of poverty in the rural eastern India, in states such as Assam, Bihar, Odisha and West Bengal. In each of these states, more than 50 per cent are classified as poor. Third, most areas experienced some poverty alleviation over the five-year period but the reduction is substantially more modest than the one suggested by the official measures. For 2004–05, the year for which there are three sets of estimates, ours are closer to the current official methodology than the one in use at that time.

TABLE 2: Poverty headcounts

	2004–05				**2009–10**			
	Rural		**Urban**		**Rural**		**Urban**	
	Engel	*IPC*	*Engel*	*IPC*	*Engel*	*IPC*	*Engel*	*IPC*
	(1)	(2)	(4)	(5)	(7)	(8)	(9)	(10)
Andhra Pradesh	32.0	32.3	20.2	23.4	27.7	22.7	16.0	17.7
Assam	85.8	36.2	45.4	21.8	73.7	39.9	43.6	25.9
Bihar	82.3	55.7	59.9	43.7	76.9	55.3	64.9	39.4
Chhattisgarh	29.0	55.1	35.4	28.4	17.3	56.1	12.5	23.6
Gujarat	37.4	39.1	19.9	20.1	46.5	26.6	21.6	17.7
Haryana	11.6	24.8	15.2	22.4	26.4	18.6	28.2	23.0
Jharkhand	78.1	51.8	36.7	23.8	40.1	41.3	47.1	31.0
Karnataka	23.4	37.5	18.5	25.9	34.5	26.1	16.9	19.5
Kerala	8.3	20.2	23.1	18.4	3.5	12.0	18.6	12.1
Madhya Pradesh	17.4	53.6	25.7	35.1	19.0	42.0	16.3	22.8
Maharashtra	20.3	47.9	15.9	25.6	20.2	29.5	15.4	18.3
Odisha	63.0	60.8	46.2	37.6	53.0	39.2	41.3	25.9
Punjab	4.6	22.1	9.0	18.7	3.1	14.6	17.5	18.1
Rajasthan	30.1	35.8	28.7	29.7	33.9	26.4	26.0	19.9
Tamil Nadu	46.3	37.5	17.4	19.7	20.5	21.2	18.2	12.7
Uttar Pradesh	27.9	42.7	37.8	34.1	32.3	39.3	39.7	31.7
West Bengal	66.8	38.3	40.0	24.4	70.3	28.8	37.3	21.9
All India	39.7	41.8	25.6	25.7	37.7	33.3	24.8	20.9

Note: The all-India rates are weighted averages of the state-level poverty headcounts, using the NSS population multipliers.

We have seen that the Engel method yields prices and corresponding poverty rates that differ–substantially for some states–from the official measures. Is there any reason to believe that the Engel poverty numbers are more credible than the official ones? In this section we present the validation exercises that we conduct – exercises that lead us to have confidence in the validity of the results revealed by the Engel method.

Our first and most elaborate exercise involves a comparison of the behavior of households we estimate to be poor with those classified as such by official lines. We do this by examining the sources from which households get their calories. An adequate intake of calories and nutrition is central to any notion of subsistence, which is why calorie norms were used to define Indian poverty lines in the 1970s. Poor families are likely to rely on cheap calories. With increases in income, families are likely to substitute away from these towards calories with better taste or status attributes (Behrman and Deolalikar, 1988). Jensen and Miller (2010) formalize this intuitive idea within a theoretical consumer choice framework and find that the evidence supports it.

We are able to compute the caloric intake of each food item consumed by a household in the NSS by multiplying the quantity consumed by the corresponding calorie conversion factor from the NSS.[17] In Appendix B we compute the average price per calorie for the main food groups reported in the NSS data. Cereals are by far the cheapest source of calories. We also plot the share of total calories from cereals versus the logarithm of total expenditure. Not surprisingly, we find a monotonic negative relationship between cereal shares and the log of total expenditure.

We use this negative relationship between cereal-calorie shares and income to evaluate the Engel-based and official poverty counts. We do this by examining the cereal-calorie shares of households in a symmetric five per cent interval around the two sets of poverty lines. If the state-wise poverty lines represent the same real expenditure level across states, one would expect these households to have similar cereal-calorie shares, despite the fact that their nominal expenditure levels vary. This hypothesis is investigated in Figure 1. Because the figure is based on households within a limited range of the expenditure distribution, we restrict the analysis to the 12 states with the largest numbers of rural households in the NSS data in 2004–05. This yields a sample of 2000 rural households in 2004–05.[18]

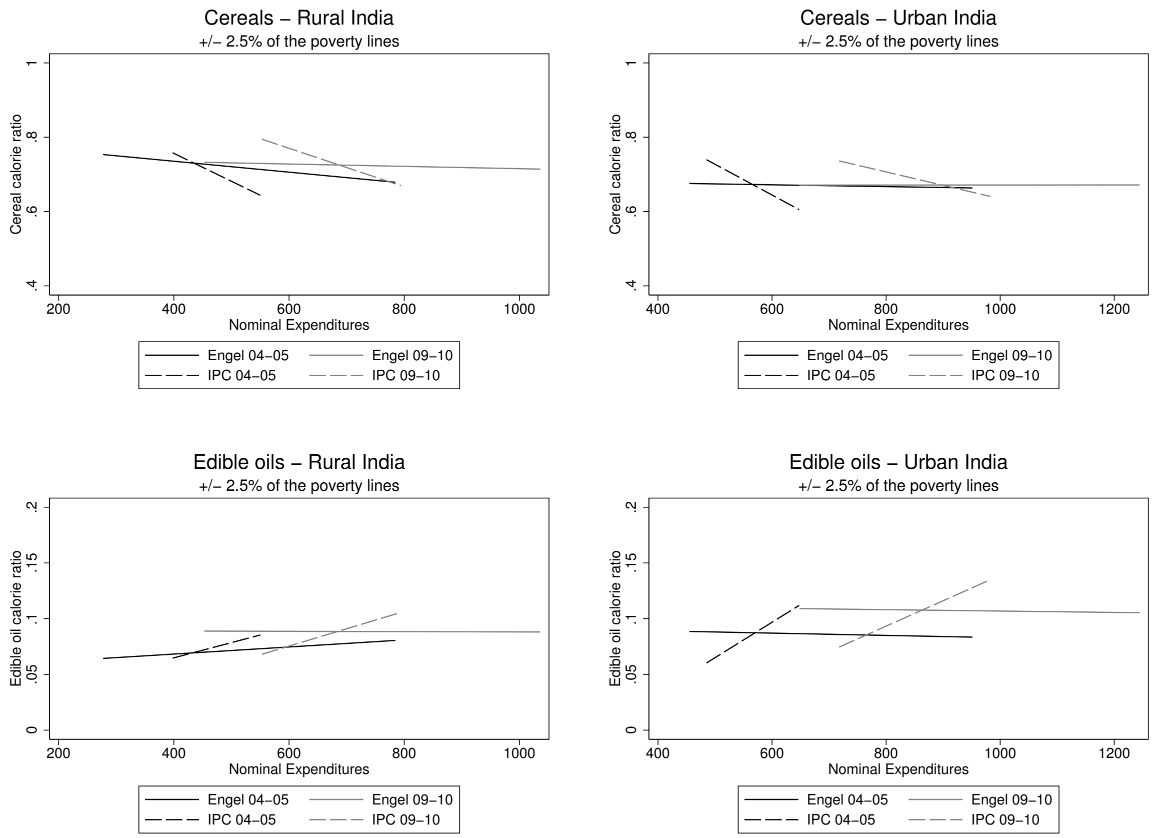
The top panel in the figure displays fitted lines for cereal-calorie shares against nominal expenditures for rural and urban sectors. Looking at the fitted lines representing families close to the Engel poverty lines we see that they are almost horizontal. In other words, households around the estimated lines in the 12 states seem to behave as if they were equally poor. Interestingly, households from states such as Assam, Bihar, West Bengal and Odisha, which have relatively high nominal poverty lines by no means diverge from the other households. The figure also graphs corresponding fitted lines for families around the official poverty lines. These households do not seem to behave as if they were equally poor. In particular, based on their higher cereal shares, households from Assam, Bihar, West Bengal and Odisha seem to act as if they were poorer than households close to the poverty lines in other states.[19] This suggests that the official methods fail to capture real cost-of-living differences across Indian states. In Appendix B we conduct a semi-parametric analysis, which indicates that these findings are neither driven by differences in

[17] These widely-used factors are based on work by the National Institute of Nutrition (Gopalan *et al.*, 1971).

[18] For consistency, we use the same 12 states for the urban sector. With a few exceptions, these states also have the largest numbers of urban households. Our rural sample in 2009–10 is 1400, mainly because of a smaller overall sample. The urban sample in the two years consists of 860 and 630 households respectively.

[19] All the slope coefficients for the official poverty lines are significantly different from zero and significantly steeper than the ones for the Engel poverty lines. None of the slope coefficients for the Engel poverty lines, except the one for rural 2004–05, are significantly different from zero.

Figure 1: Calorie shares and nominal expenditure levels



Note: The graphs in the figure display simple fitted lines using only observations on households with expenditure levels that are 2.5 per cent above and below the relevant poverty line.

relative food prices or other observed household characteristics nor by functional form assumptions. It is also reassuring that the shares seem to be stable over time and hence, our poverty lines seem to identify the same real income level in the two periods. The cereal shares around the official poverty lines are however higher in the last time period, which indicates that these lines on average represent a lower real income level compared to the official lines in the first period.[20]

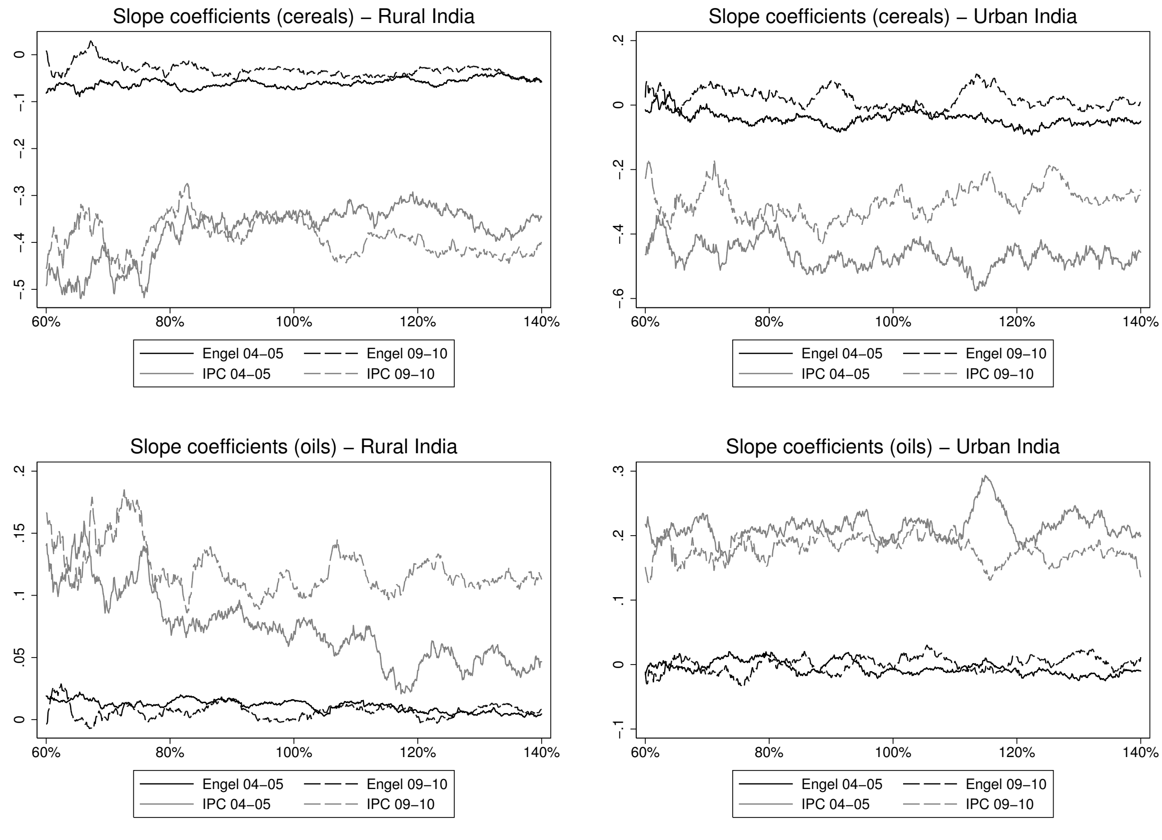
As an alternative to cereals, we could use a commodity which is consumed by most households across the country and whose share *increases* monotonically with real income. The food group edible oils & fats is such a commodity. In Appendix B we show that there is a positive relationship between edible oil-calorie shares and the logarithm of total expenditure, although this relationship is weaker than the one for cereals. In the bottom panel of Figure 1, we show fitted lines corresponding to those in the top panel but using edible oils instead of cereals. Once again, the Engel estimates provide no indication of any systematic differences across states while around

[20]We have conducted t-tests to check this more formally, and we are unable to reject a null hypothesis stating that the shares for households close to the Engel poverty lines in the two years are the same (p-value=0.337 for rural and p-value=0.758 for urban), whereas we are able to reject such a null for the official poverty lines (p-value<0.001 for both urban and rural areas).

the official lines, oil-calorie shares are rising in nominal income.[21]

The above analysis has used the set of poverty lines derived earlier in this section. In principle however, our estimated prices should provide us with comparable households across states for any interval in the distribution of real expenditure. To see whether the above pattern is robust to alternative poverty lines, we scale the all-India poverty line up and down and for each multiple of the original poverty line we estimate the linear relationship between calorie shares for cereals and nominal income for households in the five per cent band around the line. We repeat this for edible oils. Figure 2 plots these slope coefficients for both the Engel and the official methods for different multiples of the original line. The 100 per cent value corresponds to the slopes in Figure 1.

FIGURE 2: Slope coefficients calorie analysis for different poverty lines



Note: The horizontal axis displays percentage of the original all-India poverty line.

For all scalar multiples of the poverty lines we use, the slopes are roughly zero for the Engel lines for both cereals and oils. This is reassuring both for our estimates of the current pattern of poverty but also as validations of this procedure for future poverty lines, which may rely on a

[21] All the oil-slope coefficients for the official poverty lines are significantly different from zero and significantly steeper than the ones for the Engel poverty lines. And again, none of the slope coefficients for the Engel poverty lines, except the one for rural 2004–05, are significantly different from zero.

definition of subsistence at a higher level. For the official lines, the slope coefficients are negative for cereals-calorie shares and positive for oil-calorie shares. This suggests that those around official lines with higher nominal incomes also have higher real incomes.

As a second and more minor validity check of our estimates we investigate the correlation between the rural and urban price indices. If, as is generally believed, markets are fairly well integrated within states, we would expect to see a substantial positive correlation in these prices and states with a high price level relative to the all-India average in one (urban or rural) sector should also have a relatively high price level in the other sector (Deaton and Tarozzi, 2000). The Engel indices do exhibit this strong correlation between rural and urban areas, with correlation coefficients of 0.92 and 0.83 in 2004–05 and 2009–10, respectively. The corresponding correlation coefficients for the implicit Planning Commission prices are also positive, but somewhat lower at 0.81 and 0.72 for these two years. A striking contrast is found in the official estimates that were used up until 2011. These measures exhibit a *negative* correlation between spatial prices in rural and urban areas (-0.34 in 2004–05). This seems implausible and suggests that the price measure in use until recently were seriously out of date.[22]

In our final validity check we use responses on household perceptions of hunger. Although hunger is not (necessarily) the same as poverty, we would expect the two measures to be positively and significantly correlated. In the NSS survey, respondents are asked whether every member of the household gets "enough food every day".[23] This is a self-reported measure of hunger and should be interpreted with the usual caveats. However, we have little reason to expect any systematic errors across states.[24] Figure 3 shows the proportion of all households reporting a lack of food.[25] These numbers are plotted against two sets of headcount ratios: those from the Engel analysis and the official poverty rates. The graphs reveal that four of the five states with the highest levels of self-reported hunger are Assam, Bihar, Odisha and West Bengal; states for which the Engel methodology also reports high poverty rates (and higher than the official numbers).

Table 3 shows the overall correlation between headcounts and self-reported hunger ratios. The poverty counts based on the official methodology are positively correlated with the self-reported hunger but the correlations are smaller than those for the Engel counts.

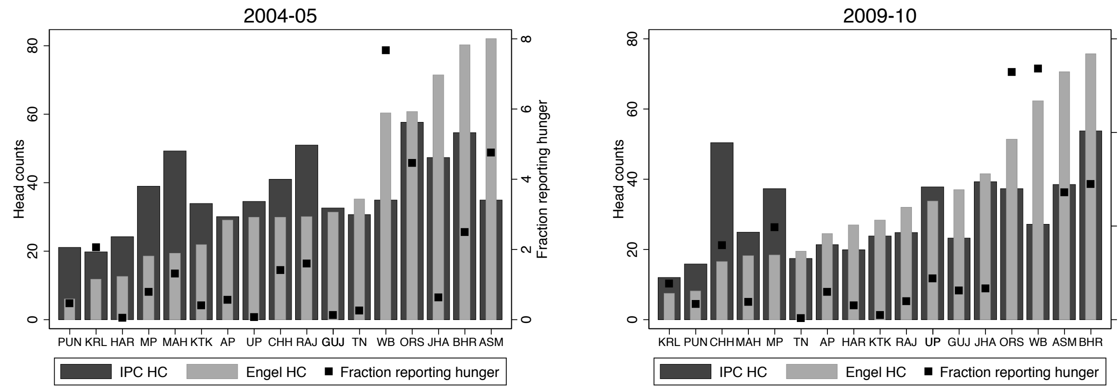
[22]See Deaton and Tarozzi (2000) and Deaton (2003) for similar findings for earlier years.

[23]These proportions are taken from the "type 2" NSS survey, because the question does not appear in the "type 1" survey that we use for the rest of our data.

[24]See Deaton and Tarozzi (2000) for a critical evaluation of this subjective measure.

[25]We combine the rural and urban headcounts, using population weights.

Figure 3: Headcounts and self-reported hunger



Note: The hunger questions from the 61st and 66th NSS surveys are not entirely consistent with each other. In NSS61, respondents are asked "Do all members of your household get enough food every day?", and are asked to choose between: "yes: every month of the year"; "some months of the year"; and "no: no month of the year". In NSS66, respondents are asked "Do all members of your household get two square meals every day?, and are asked to choose between: "yes: every month of the year"; "some months of the year"; and "no: no month of the year". This discrepancy could explain the relatively large drop in the number of households reporting hunger over time. However, the discrepancy is not a major concern because we do not compare households between survey rounds.

Table 3: Correlations between self-reported hunger and headcounts

	Engel		**IPC**	
	Rural	Urban	Rural	Urban
	(1)	(2)	(3)	(4)
2004–05	0.55	0.74	0.20	0.28
2009–10	0.64	0.53	0.41	0.41

Note: The table presents correlations between state-wise price levels in the rural and the urban sector.

5 Robustness analysis

Even though the validation of our results makes us confident that the Engel method provides sensible measures of price levels and poverty, we still want to perform some robustness analysis relating to our empirical specification and sample. We start by including relative food and non-food prices as an additional control since these may influence the budget share for food. We then limit ourselves to look at the intertemporal price movements and estimate our model separately for each state addressing the potential worry that tastes may differ across geographical regions (see Atkin, 2013, for a discussion of this). This identifies price changes over the five year period for each state and we compare these with the inter-temporal estimates derived in Section 4. We also investigate the assumption of a log-linear functional form. We do this by estimating the Engel curves semi-parametrically for each state and also by comparing our estimates with those from a more flexible quadratic demand specification. All the estimates presented above are based on households with exactly two children and two adults. As a robustness analysis we re-estimate our model using all available households and find very similar results. Further, one might worry that noise in the expenditure variable could downwardly bias the coefficient for the logarithm of total expenditure, as the variable appears on both sides of Equation (1). We address this concern in a final robustness check, using the logarithm of the village level mean as an instrument for the logarithm of total expenditure. Details on each of these checks are given below.

Turning to our first specification check, it is possible that the budget shares are influenced by relative prices. We explore this by including the ratio of food and non-food prices as an additional control variable in our Engel estimation.[26] This ratio is constructed from unit values obtained by simply dividing expenditures by the quantity consumed for items for which both these are available. This is the case for 127 food items and 41 non-food items. We use median unit values for all these 168 consumption items at the district level.[27] Although unit values are different from prices, this should give a proxy to the relative food and non-food price relation in different locations. Details on construction of the relative price variable are given in Appendix C.

[26]When budget shares depend on relative prices, the cost-of-living in the demand system becomes income specific in that the cost-of-living comparison will depend on the income level chosen for evaluation. Hence, the Engel based method with relative prices measures the cost-of-living for one specific income level. This reference utility level need not be the same as that underlying conventional price indices. See Beatty and Crossley (2012) for a discussion of this. It is therefore reassuring that our main findings are not sensitive to including the relative prices in the estimation of cost-of-living, and hence we have no reason to expect that including relative prices and an alternative reference utility level is quantitatively important.

[27]Relative prices are used at the district level, as a state specific relative price variable would make the identification through the dummies impossible. We use the median rather than the mean because it is less sensitive to outliers.

With the relative price control, the Engel curve in the demand system is given by:

$$m_{hdst} = a + b(\ln y_{hdst} - \ln P_{st}) + \gamma(\ln P^f_{dst} - \ln P^n_{dst}) + \theta X_{hdst} + \varepsilon_{hdst}, \tag{5}$$

where P^f_{dst} is the price of food and P^n_{dst} is the price of non-food items in district d in state s at time t. The only unknown variable in this regression is as before, the overall state price level P_{st} and the identification of this is as before. Table A.3 presents the estimated parameters, and Table A.4 and A.5 display the corresponding price estimates. We can see that these estimated parameters imply almost identical price estimates as those presented in our main analysis.

We next compare our inter-temporal price changes from our pooled model in Equation (2), with estimates of the same changes from estimating the model separately for each state and rural and urban sectors. By normalizing the price level in the first period for each state and urban and rural sector to unity we can pick up the price level in the second period by estimating:

$$m_{ht} = a + b(\ln y_{ht}) + \theta X_{ht} + dD_{t+1} + \varepsilon_{ht}, \tag{6}$$

and using the dummy-coefficient, d to compute:

$$P_{t+1} = e^{-\frac{d}{b}}. \tag{7}$$

The third row in Table A.6 presents the overall price estimates for the rural and the urban sector. It is comforting that this disaggregated analysis gives almost identical state-wise price changes as our pooled results.

As a specification check we relax the assumption of a log-linear relationship between the budget share for food and total expenditures. We first present estimates from a semi-parametric kernel analysis. The analysis is based on removing the effects of all our covariates in Equation (2) other than the logarithm of nominal expenditures, using differencing. The resulting residuals are plotted against the logarithm of nominal expenditures in Figure A.1, separately for each of the major states and time periods. While this procedure forces the partial effects of the covariates to be linear and similar over time and between states, the effect of the log of expenditure is allowed to have a more flexible functional form and to vary across states. We find that the plotted lines are close to being log-linear and there is little variation, both over time and between states. Hence, the kernel analysis suggests that our main results are not driven by our functional form assumptions.

As a further check on functional form, we estimate the following quadratic demand system (Banks *et al.*, 1997; Dickens *et al.*, 1993):

$$m_{hst} = a + b_1(\ln y_{hst} - \ln P_{st}) + b_2(\ln y_{hst} - \ln P_{st})^2 + \theta X_{hst} + \varepsilon_{hst}. \tag{8}$$

The overall price component, P_{st}, is identified directly using non-linear iteration and state- and time-specific dummy variables. For both urban and rural sectors, the coefficients for the squared expenditure terms are statistically significant but small. The other coefficients are comparable with those from the linear specification.

The third column of Table A.4 and A.5 reports the corresponding spatial price measures. These confirm, and strengthen, our first two findings. There is more price dispersion across states than implied by the official measures, and the price indices indicate a relatively high cost-of-living in the eastern states. The fourth row of Table A.6 reports the implied inter-temporal price measures. These are very similar to those from our main specification.

In Section 4, we restricted our sample to households with two children and two adults. Estimates from the full sample of households generate similar expenditure elasticities for the urban sector and slightly larger elasticities for the rural sector.[28] This holds whether we use per capita expenditure in the Engel regression, or whether we adjust household expenditure using OECD's equivalence scale.[29] The corresponding spatial prices are presented in Column (4) and (5) of Table A.4 and A.5. The geographical pattern of prices is similar. Price changes are slightly lower than with our restricted sample but indicate, as before, higher cost-of-living increases than suggested by the official measures. Thus, our sampling restrictions do not drive our main findings.

As a final robustness check we use the logarithm of the mean expenditure at the village level as an instrument for the logarithm of expenditure at the household level.[30] It can be shown that this gives consistent estimates of the spatial price levels even if both food expenditure and total expenditure are measured with noise (see Appendix D for details). As shown in the last column in Table A.4 and A.5, this gives very similar spatial patterns as in the main specification. We also see from Table A.6 that the intertemporal price changes estimated by the IV-strategy are almost identical to those from the main specification. We therefore conclude that our findings are not driven by pure noise in reported expenditure.

[28] The extended sample's expenditure elasticity of +0.79 versus +0.77 from the main specification.

[29] This scale gives a weight of 1 to the first adult, a weight of 0.7 to the rest of the adults in the household, and a weight of 0.5 to all children. We define a child as a person of age below 16 years old.

[30] The villages usually have 8-10 sampled households. Instruments based on the district level means give very similar resultants as the ones presented here.

6 Conclusion

In this paper, we have proposed a method for poverty comparisons in which price levels and poverty lines are estimated based on observed consumer behavioural and the assumption that equally poor households behave in the same way along specific dimensions. We estimate Engel curves for food and use those to reveal state specific price levels and poverty lines. We validate the resulting price and poverty findings using alternative methods. The Engel approach has been used in several contexts to analyze potential biases in prices. We apply it for the first time in estimating both spatial and inter-temporal variation in prices and corresponding poverty counts, and we are the first to conduct an extensive validation of the estimates from the method.

Our poverty estimates differ in significant ways from those published by the Indian Planning Commission. We find much higher spatial variation in prices and poverty across Indian states. The divergence from official poverty rates follows a specific regional pattern. Poverty in the rural areas of the eastern states of Assam, Bihar, Odisha and West Bengal is consistently higher than official figures and exceeds 50 per cent in both survey years. We also find that the decrease in overall poverty over our five-year period is much more modest than what is suggested by official statistics. All these findings are robust to a variety of robustness checks.

Given these different estimates, it is particularly important to ask whether our estimates are credible. The methods we use to examine the validity of our estimates are an important methodological and empirical contribution of our paper. Our most elaborate validation check is to study the consumption behavior of households in a narrow band around our state poverty lines. We find that these households consume similar shares of their calories from different food groups, suggesting that they have the same real income, even though their nominal income varies because prices differ across states. This is reassuring for our price and poverty estimates. Future research will reveal whether the Engel method performs equally well also in other applications.

References

Almås, I. (2012). International Income Inequality: Measuring PPP Bias by Estimating Engel Curves for Food. *American Economic Review*, **102**(1), 1093–1117.

Almås, I. and Johnsen, Å. (2013). The cost of living in china: Implications for inequality and poverty. *Memorandum, Department of Economics, University of Oslo*, **06/2013**.

Almås, I. and Sørensen, E. (2012). Global Income Inequality and Cost-of-Living Adjustment: The Geary-Allen World Accounts. *Department of Economics, NHH, Discussion paper*, **20**.

Atkin, D. (2013). Trade, Tastes, and Nutrition in India. *The American Economic Review*, **103**(5), 1629–1663.

Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic Engel Curves and Consumer Demand. *Review of Economics and Statistics*, **79**(4), 527–539.

Barrett, G. and Brzozowski, M. (2010). Using Engel Curves to Estimate the Bias in the Australian CPI. *The Economic Record, The Economic Society of Australia*, **86**(272), 1–14.

Beatty, T. and Crossley, T. (2012). Lost in translation: What do Engel Curves tell us about the cost of living? *Working Paper, University of Minnesota.*

Beatty, T. and Larsen, E. (2005). Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index. *Canadian Journal of Economics/Revue canadienne d'économique*, **38**(2), 482–499.

Behrman, J. and Deolalikar, A. (1988). Health and Nutrition. *Handbook of Development Economics*, **1**, 631–711.

Blundell, R., Duncan, A., and Pendakur, K. (1998). Semiparametric Estimation and Consumer Demand. *Journal of Applied Econometrics*, **13**(5), 435–461.

Carvalho Filho, I. and Chamon, M. (2006). *The myth of post-reform income stagnation in Brazil*. International Monetary Fund.

Chattopadhyay, S. (2010). District Level Poverty Estimation: A Spatial Approach. *Economics Bulletin*, **30**(4), 2962–2977.

Chung, C., Gibson, J., and Kim, B. (2010). CPI Mismeasurements and Their Impacts on Economic Management in Korea. *Asian Economic Papers*, **9**(1), 1–15.

Coondoo, D., Majumder, A., and Chattopadhyay, S. (2011). Estimating Spatial Consumer Price Indices Through Engel Curve Analysis. *Review of Income and Wealth*, **57**(1), 138–155.

Costa, D. (2001). Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves. *Journal of Political Economy*, **109**(6), 1288–1310.

Deaton, A. (1988). Quality, Quantity, and Spatial Variation of Price. *The American Economic Review*, **78**(3), 418–430.

Deaton, A. (2003). Prices and Poverty in India, 1987-2000. *Economic & Political Weekly*, **38**(4), 362–368.

Deaton, A. (2008). Price Trends in India and Their Implications for Measuring Poverty. *Economic & Political Weekly*, **43**(6), 43–49.

Deaton, A. (2010). Price Indexes, Inequality, and The Measurement of World Poverty. *The American Economic Review*, **100**(1), 1–34.

Deaton, A. and Dupriez, O. (2011). Spatial price differences within large countries. Technical report.

Deaton, A. and Muellbauer, J. (1980a). An Almost Ideal Demand System. *The American Economic Review*, **70**(3), 312–326.

Deaton, A. and Muellbauer, J. (1980b). *Economics and Consumer Behavior*. Cambridge University Press.

Deaton, A. and Tarozzi, A. (2000). Prices and Poverty in India. Manuscript.

Dickens, R., Fry, V., and Pashardes, P. (1993). Non-linearities and Equivalence Scales. *The Economic Journal*, **103**(417), 359–368.

Diewert, E. (1978). Superlative Index Numbers and Consistency in Aggregation. *Econometrica*, **46**(4), 883–900.

Engel, E. (1857). Die productions und consumtions verhältnisse des königreichs sachsen. *Bulletin de lInstitut International de la Statistique*, (9).

Engel, E. (1895). *Die Lebenskosten belgischer Arbeiter-Familien früher und jetzt*. C. Heinrich.

Gibson, J., Stillman, S., and Le, T. (2008). CPI Bias and Real Living Standards in Russia During the Transition. *Journal of Development Economics*, **87**(1), 140–160.

Gopalan, C., Sastri, B., and Balasubramanian, S. (1971). Nutritive Values of Indian Foods. *Hyderabad, National Institute of Nutrition*.

Government of India (1979). Report of The Task Force on Projections of Minimum Need and Effective Consumption Demand.

Government of India (1993). Report of the Expert Group on Estimation of the Proportion and Number of Poor.

Government of India (2009). Report of the Expert Group to Review the Methodology for Estimation of Poverty.

Government of India (2011). Press Note on Poverty Estimates.

Hamilton, B. (2001). Using Engel's Law to Estimate CPI Bias. *The American Economic Review*, **91**(3), 619–630.

Hill, R. J. (2004). Constructing price indexes across space and time: the case of the european union. *American Economic Review*, pages 1379–1410.

Jensen, R. and Miller, N. (2010). A Revealed Preference Approach to Measuring Hunger and Undernutrition. *NBER Working Paper*, (16555).

Larsen, E. (2007). Does the CPI Mirror the Cost of Living? Engel's Law Suggests not in Norway. *The Scandinavian Journal of Economics*, **109**(1), 177–195.

Leser, C. (1963). Forms of Engel Functions. *Econometrica*, **31**(4), 694–703.

Mishra, A. and Ray, R. (2011). Prices, Inequality, and Poverty: Methodology and Indian Evidence. *Review of Income and Wealth*, **57**(3), 428–448.

Nakamura, Emi, J. S. and Liu, M. (2013). Is china slowing? engel curve estimates of chinese inflation and growth. *Working paper (unpublished)*.

Neary, J. P. (2004). Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income. *American Economic Review*, **94**(5), 1411–1428.

Olivia, S. and Gibson, J. (2012). Using Engel Curves to Measure CPI Bias for Indonesia. *Monash Econometrics and Business Statistics Working Papers*, **13**.

Pendakur, K. (2002). Taking prices seriously in the measurement of inequality. *Journal of Public Economics*, **86**(1), 47–69.

Pollak, R. A. and Wales, T. J. (1981). Demographic Variables in Demand Analysis. *Econometrica*, **49**(6), 1533–1551.

Rao, P. (1990). The Estimation of the Mean Squared Error of Small-area Estimators. *Journal of the American statistical association*, **85**(409), 163–171.

Ravallion, M. (2015). Toward better global poverty measures. *Center for Global Development Working Paper*, (417).

Srinivasan, T. (2007). Poverty Lines in India: Reflections After the Patna Conference. *Economic & Political Weekly*, **42**(41), 4155–4165.

Subramanian, S. (2011). The Poverty Line: Getting It Wrong Again. *Economic & Political Weekly*, **46**(48), 66–70.

Working, H. (1943). Statistical Laws of Family Expenditure. *Journal of the American Statistical Association*, **38**(221), 43–56.

Yatchew, A. (2003). *Semiparametric regression for the applied econometrician*. Cambridge University Press.

Appendix A Tables and figures

Table A.1: Household descriptive statistics

	Rural				Urban			
	2004–05		**2009–10**		**2004–05**		**2009–10**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Monthly per capita expenditure (Rs.)	578.90	409.64	952.51	723.59	1101.63	923.59	1851.60	1788.62
Children in HH (#)	2.48	1.87	2.17	1.71	1.88	1.68	1.68	1.55
Adults in HH (#)	3.62	1.88	3.61	1.82	3.72	1.95	3.66	1.89
HH head's age (years)	46.06	13.23	46.62	12.96	46.14	13.14	46.38	13.30
Proportion of:								
Females in household	0.49	0.16	0.48	0.16	0.48	0.18	0.48	0.18
Self-employed (non-agriculture)	0.17	0.37	0.16	0.37				
Self-employed (agriculture)	0.39	0.49	0.35	0.48				
Agricultural labour	0.25	0.43	0.25	0.43				
Self-employed					0.43	0.50	0.42	0.49
Salaried labour					0.39	0.49	0.37	0.48
Casual labour					0.12	0.32	0.14	0.35
Hindu	0.84	0.37	0.84	0.37	0.78	0.42	0.78	0.41
Muslim	0.11	0.32	0.12	0.32	0.16	0.37	0.16	0.37
Christian	0.02	0.14	0.02	0.15	0.02	0.15	0.03	0.16
Other religion	0.03	0.17	0.03	0.16	0.04	0.19	0.04	0.19
Cultivated land (hectares)								
None	0.36	0.48	0.40	0.49	0.90	0.29	0.91	0.28
Less than 2	0.51	0.50	0.48	0.50	0.08	0.26	0.07	0.25
2 or more	0.13	0.34	0.12	0.32	0.02	0.14	0.02	0.14
Free meals outside household (#)	0.08	0.18	0.09	0.17	0.06	0.19	0.06	0.18
Households (# 1000)	78.64		58.61		44.40		40.91	

Note: These summary statistics are for our sample of 30 Indian states and union territories. All variables are weighted by the population multipliers provided by the NSS.

Table A.2: Price indices

	Spatial price indices								Normalized to all-India 2004–05			
	2004–05				2009–10				2009–10			
	Rural		Urban		Rural		Urban		Rural		Urban	
	Engel	*IPC*	*Engel*	*IPC*	*Engel*	*IPC*	*Engel*	*IPC*	*Engel*	*IPC*	*Engel*	*IPC*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Andhra Pradesh	96.8	96.0	91.1	96.6	102.6	102.0	97.0	105.8	163.7	153.6	154.5	159.0
Assam	171.2	105.8	160.5	103.0	141.8	101.7	132.0	99.5	226.2	153.2	210.1	149.5
Bihar	126.5	96.0	120.0	90.3	117.1	96.4	115.7	88.6	186.8	145.2	184.3	133.1
Chhattisgarh	71.0	88.3	96.8	88.2	62.6	90.8	67.5	92.1	100.0	136.7	107.5	138.4
Gujarat	109.6	111.1	113.3	113.1	122.5	106.7	110.3	108.7	195.5	160.7	175.6	163.3
Haryana	97.9	117.2	94.0	107.5	123.4	116.4	114.3	111.4	197.0	175.3	181.9	167.4
Jharkhand	117.0	89.6	117.5	91.2	85.1	90.6	113.6	94.9	135.8	136.5	180.9	142.6
Karnataka	82.7	92.5	89.0	100.9	95.3	92.6	93.2	103.7	152.1	139.4	148.4	155.8
Kerala	93.5	119.0	111.1	100.3	82.0	114.0	100.8	94.9	130.8	171.7	160.4	142.6
Madhya Pradesh	63.8	90.4	80.7	91.3	65.2	92.9	72.2	88.1	104.0	139.9	114.9	132.4
Maharashtra	78.6	107.4	88.4	108.4	92.6	109.4	98.1	109.8	147.7	164.7	156.1	164.9
Odisha	93.3	90.3	99.8	85.3	91.5	83.4	99.4	84.1	146.0	125.6	158.2	126.3
Punjab	88.8	120.4	92.7	110.3	91.9	122.1	102.4	109.7	146.7	183.8	163.0	164.9
Rajasthan	100.6	105.8	96.7	97.5	114.4	111.0	99.9	96.6	182.6	167.2	159.1	145.2
Tamil Nadu	107.6	97.8	92.1	96.1	88.7	94.0	96.3	91.5	141.6	141.5	153.4	137.4
Uttar Pradesh	83.9	96.4	98.0	91.3	87.8	97.6	97.4	91.4	140.1	147.0	155.0	137.3
West Bengal	133.1	98.6	125.0	98.3	129.0	94.6	113.6	94.9	205.8	142.4	180.9	142.5
All India	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	161.4	150.6	160.6	150.3
CV	0.22	0.15	0.20	0.12	0.21	0.14	0.18	0.11				

Note: The table displays the different price level estimates. *Engel* indicates that the estimates are from the Engel method, whereas *IPC* indicates that the estimates are based on the official methods. Standard errors for the Engel based price estimates are reported in Table A.4 and A.5. CV denotes the coefficient of variation. The all-India values are population-weighted averages of state-level prices, normalized to 100.

Table A.3: Regressions for robustness checks

Dep. var.: Budget share for food (%)	**Engel$_{\mathbf{rp}}$** (1)	**QEngel** (2)	**Engel$_{\mathbf{fs}}$** (3)	**Engel$_{\mathbf{fs\text{-}eq}}$** (4)	**Engel$_{\mathbf{IV}}$** (5)
Rural					
Log of household expenditure	-12.628	38.113	-11.694	-11.766	-10.229
	(0.227)	(4.116)	(0.107)	(0.110)	(0.480)
Log of household expenditure squared		-3.219			
		(0.267)			
Log of relative food/non-food prices	-0.085				
	(0.719)				
Observations	14258	14258	137152	137152	14258
R^2	0.379	0.395	0.357	0.359	0.371
Urban					
Log of household expenditure	-13.761	6.168	-13.115	-13.321	-15.498
	(0.191)	(4.101)	(0.112)	(0.116)	(0.428)
Log of household expenditure squared		-1.222			
		(0.249)			
Log of relative food/non-food prices	1.902				
	(0.786)				
Observations	9113	9113	85300	85300	9113
R^2	0.523	0.524	0.459	0.469	0.516

Note: Robust standard errors, clustered at the NSS first stage sample unit, are given in parentheses. The subscript rp denotes relative price controls, fs denotes that the full sample is used in the estimation, fs-eq denotes full sample with use of equivalence scaling, and IV denotes that the instrument variable approach is used. Controls that are included but not shown in the table are: the age of the household head; the proportion of females in the household; three occupation dummies for each sector (urban and rural); three religion dummies; the number of free meals taken outside the home; and dummies for cultivated land. In addition, Engel$_{fs}$ includes controls for the number of children and the number of adults, and their squares.

Table A.4: Robustness checks: Spatial prices rural

	Engel		**Engel$_{rp}$**		**QEngel**		**Engel$_{fs}$**		**Engel$_{fs\text{-}eq}$**		**Engel$_{IV}$**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2004–05												
Andhra Pradesh	96.8	(2.5)	96.9	(2.6)	92.2	(2.8)	99.8	(1.8)	100.0	(1.7)	93.9	(3.)2
Assam	171.2	(7.2)	171.4	(7.2)	186.7	(12.0)	175.9	(4.5)	176.3	(4.5)	192.4	(10.9)
Bihar	126.5	(4.0)	126.3	(4.0)	133.5	(6.5)	127.7	(1.9)	127.2	(1.9)	137.8	(5.8)
Chhattisgarh	71.0	(3.5)	71.0	(3.5)	68.1	(3.3)	78.3	(2.3)	77.8	(2.3)	68.4	(4.3)
Gujarat	109.6	(4.5)	109.7	(4.5)	104.3	(4.7)	112.2	(2.7)	112.1	(2.7)	105.7	(5.5)
Haryana	97.9	(6.3)	97.9	(6.3)	99.3	(5.8)	99.3	(3.2)	99.8	(3.2)	88.9	(7.3)
Jharkhand	117.0	(5.0)	117.0	(5.0)	132.8	(8.4)	117.0	(2.5)	116.0	(2.4)	128.5	(7.2)
Karnataka	82.7	(3.0)	82.7	(3.0)	78.8	(2.9)	84.2	(1.8)	84.6	(1.8)	78.6	(3.6)
Kerala	93.5	(4.0)	93.5	(4.0)	94.2	(3.6)	88.9	(2.2)	90.5	(2.2)	82.0	(5.1)
Madhya Pradesh	63.8	(2.9)	63.8	(2.9)	62.7	(2.8)	61.5	(1.6)	61.3	(1.6)	59.2	(3.5)
Maharashtra	78.6	(2.5)	78.7	(2.6)	76.6	(2.4)	72.4	(1.2)	72.7	(1.2)	72.5	(3.3)
Orissa	93.3	(4.1)	93.3	(4.1)	98.1	(5.5)	100.1	(2.3)	99.9	(2.3)	98.6	(5.3)
Punjab	88.8	(4.8)	88.7	(4.9)	89.0	(4.3)	87.5	(2.5)	88.3	(2.5)	78.2	(5.8)
Rajasthan	100.6	(4.8)	100.6	(4.8)	97.9	(5.1)	99.4	(2.3)	99.1	(2.3)	98.4	(5.7)
Tamil Nadu	107.6	(3.9)	107.7	(4.1)	105.9	(4.5)	99.2	(2.3)	99.9	(2.3)	107.4	(4.9)
Uttar Pradesh	83.9	(2.9)	83.8	(2.9)	80.8	(3.0)	83.7	(1.2)	83.2	(1.1)	80.4	(3.5)
West Bengal	133.1	(3.8)	133.1	(3.8)	135.5	(6.1)	136.3	(2.3)	136.7	(2.3)	142.1	(5.5)
All India	100.0		100.0		100.0		100.0		100.0		100.0	
CV	0.22		0.22		0.24		0.24		0.24		0.27	
2009–10												
Andhra Pradesh	102.6	(3.3)	102.6	(3.3)	101.2	(3.5)	105.0	(2.3)	105.2	(2.3)	100.2	(4.0)
Assam	141.8	(7.4)	141.9	(7.5)	152.7	(11.7)	158.6	(4.8)	159.1	(4.8)	157.4	(10.5)
Bihar	117.1	(4.6)	117.0	(4.6)	122.4	(5.9)	120.2	(2.4)	119.9	(2.4)	125.4	(6.3)
Chhattisgarh	62.6	(5.1)	62.6	(5.1)	65.3	(4.8)	63.1	(2.5)	62.5	(2.5)	57.8	(5.9)
Gujarat	122.5	(5.6)	122.6	(5.7)	116.1	(6.0)	117.1	(3.0)	117.0	(3.0)	123.2	(7.1)
Haryana	123.4	(8.3)	123.5	(8.3)	124.1	(8.2)	119.3	(4.0)	119.7	(4.0)	119.2	(10.1)
Jharkhand	85.1	(4.8)	85.1	(4.8)	84.4	(5.4)	102.1	(3.3)	101.9	(3.2)	86.6	(6.1)
Karnataka	95.3	(3.9)	95.3	(3.9)	93.3	(3.9)	85.8	(2.1)	86.2	(2.1)	92.4	(4.8)
Kerala	82.0	(4.0)	82.0	(4.0)	91.8	(3.8)	76.9	(1.9)	78.1	(2.0)	68.9	(5.1)
Madhya Pradesh	65.2	(3.2)	65.2	(3.2)	66.6	(3.0)	66.9	(1.7)	66.6	(1.7)	60.3	(3.8)
Maharashtra	92.6	(3.0)	92.6	(3.0)	89.8	(2.9)	82.6	(1.5)	82.9	(1.4)	87.7	(3.8)
Orissa	91.5	(4.1)	91.5	(4.1)	94.2	(4.7)	94.4	(2.7)	94.3	(2.7)	94.9	(5.2)
Punjab	91.9	(5.9)	91.8	(5.9)	90.2	(4.9)	96.2	(3.5)	97.0	(3.5)	82.3	(6.9)
Rajasthan	114.4	(6.3)	114.5	(6.3)	111.2	(6.5)	104.9	(2.6)	104.5	(2.6)	115.6	(7.8)
Tamil Nadu	88.7	(3.3)	88.8	(3.4)	85.5	(3.1)	82.7	(2.0)	83.3	(2.0)	83.6	(4.2)
Uttar Pradesh	87.8	(3.2)	87.8	(3.2)	85.0	(3.2)	92.1	(1.4)	91.5	(1.3)	86.5	(4.0)
West Bengal	129.0	(4.3)	128.9	(4.3)	131.0	(5.9)	128.6	(2.8)	129.0	(2.8)	139.0	(6.1)
All India	100.0		100.0		100.0		100.0		100.0		100.0	
CV	0.21		0.21		0.22		0.22		0.22		0.28	

Note: Standard errors for the nonlinear price expressions are computed using the Delta method.

Table A.5: Robustness checks: Spatial prices urban

	Engel		**Engel$_{\mathbf{rp}}$**		**QEngel**		**Engel$_{\mathbf{fs}}$**		**Engel$_{\mathbf{fs\text{-}eq}}$**		**Engel$_{\mathbf{IV}}$**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2004–05												
Andhra Pradesh	91.1	(2.9)	91.7	(2.9)	90.0	(2.9)	87.9	(2.2)	88.6	(2.2)	91.0	(2.6)
Assam	160.5	(10.7)	158.5	(10.5)	162.7	(11.8)	150.8	(7.2)	149.6	(7.0)	151.2	(9.2)
Bihar	120.0	(7.5)	121.8	(7.7)	121.7	(8.3)	128.3	(3.7)	126.8	(3.6)	113.6	(6.4)
Chhattisgarh	96.8	(6.4)	96.2	(6.3)	96.5	(6.5)	83.8	(3.5)	84.0	(3.5)	95.6	(5.7)
Gujarat	113.3	(4.2)	112.7	(4.2)	112.2	(4.1)	115.9	(2.9)	115.8	(2.8)	113.6	(3.7)
Haryana	94.0	(5.1)	94.5	(5.1)	95.0	(4.9)	93.0	(2.8)	93.1	(2.7)	96.9	(4.7)
Jharkhand	117.5	(8.1)	117.8	(8.1)	121.0	(8.4)	133.2	(3.8)	131.7	(3.7)	114.0	(7.2)
Karnataka	89.0	(3.1)	89.5	(3.1)	87.9	(3.0)	87.6	(2.0)	87.8	(2.0)	90.7	(2.7)
Kerala	111.1	(5.7)	110.5	(5.6)	109.2	(5.4)	99.6	(2.8)	100.7	(2.8)	113.3	(5.1)
Madhya Pradesh	80.7	(3.1)	80.3	(3.1)	80.3	(3.0)	77.0	(1.7)	77.3	(1.7)	81.4	(2.8)
Maharashtra	88.4	(2.3)	87.6	(2.3)	88.8	(2.2)	90.5	(1.4)	90.8	(1.4)	90.8	(2.1)
Orissa	99.8	(5.4)	100.3	(5.4)	100.5	(5.7)	111.9	(4.6)	111.4	(4.5)	96.6	(4.7)
Punjab	92.7	(4.1)	95.1	(4.4)	93.4	(4.1)	91.5	(2.3)	91.5	(2.3)	94.9	(3.8)
Rajasthan	96.7	(4.7)	97.7	(4.8)	97.1	(4.7)	101.1	(2.6)	100.7	(2.5)	97.6	(4.3)
Tamil Nadu	92.1	(2.6)	90.4	(2.7)	91.5	(2.6)	95.8	(2.0)	96.7	(1.9)	92.3	(2.3)
Uttar Pradesh	98.0	(3.7)	98.5	(3.7)	98.8	(3.8)	96.3	(1.9)	95.8	(1.8)	97.1	(3.3)
West Bengal	125.0	(4.8)	125.6	(4.8)	124.2	(5.3)	121.5	(2.6)	121.2	(2.5)	120.6	(4.1)
All India	100.0		100.0		100.0		100.0		100.0		100.0	
CV	0.20		0.21		0.21		0.21		0.21		0.19	
2009–10												
Andhra Pradesh	97.0	(3.3)	98.0	(3.4)	97.1	(3.2)	99.4	(2.2)	99.9	(2.1)	98.4	(3.0)
Assam	132.0	(9.9)	129.1	(9.8)	133.4	(10.9)	128.2	(6.1)	127.6	(6.0)	125.2	(8.5)
Bihar	115.7	(6.3)	117.3	(6.5)	117.8	(6.9)	110.2	(3.3)	109.5	(3.2)	109.9	(5.5)
Chhattisgarh	67.5	(5.5)	68.1	(5.5)	70.2	(4.9)	70.9	(3.5)	70.6	(3.4)	70.2	(5.2)
Gujarat	110.3	(4.4)	108.8	(4.4)	108.8	(4.4)	110.4	(2.8)	110.4	(2.7)	110.2	(3.9)
Haryana	114.3	(5.2)	114.6	(5.3)	114.7	(5.0)	107.4	(3.3)	107.6	(3.3)	115.4	(4.8)
Jharkhand	113.6	(7.7)	114.3	(7.8)	116.1	(8.3)	119.2	(4.8)	117.8	(4.7)	109.7	(6.9)
Karnataka	93.2	(3.7)	94.2	(3.7)	91.9	(3.6)	96.6	(2.1)	96.7	(2.1)	93.6	(3.3)
Kerala	100.8	(6.3)	101.5	(6.3)	100.7	(5.9)	90.9	(2.7)	92.1	(2.7)	103.6	(5.8)
Madhya Pradesh	72.2	(3.2)	73.0	(3.3)	72.5	(3.0)	68.5	(1.6)	68.8	(1.6)	74.2	(3.0)
Maharashtra	98.1	(3.2)	97.6	(3.1)	98.3	(3.0)	99.7	(1.8)	99.9	(1.7)	99.7	(2.9)
Orissa	99.4	(6.8)	99.3	(6.9)	99.4	(6.9)	101.2	(3.8)	100.8	(3.7)	96.8	(5.9)
Punjab	102.4	(4.5)	104.2	(4.7)	101.9	(4.3)	102.9	(3.1)	102.9	(3.0)	103.1	(4.1)
Rajasthan	99.9	(5.7)	99.8	(5.7)	99.7	(5.6)	106.2	(3.0)	105.6	(2.9)	100.2	(5.1)
Tamil Nadu	96.3	(2.9)	95.1	(2.9)	95.9	(2.8)	92.3	(2.0)	93.1	(1.9)	96.6	(2.6)
Uttar Pradesh	97.4	(3.8)	97.1	(3.8)	97.1	(3.9)	96.0	(1.8)	95.5	(1.8)	96.5	(3.4)
West Bengal	113.6	(4.6)	113.6	(4.5)	114.5	(4.9)	110.7	(2.6)	110.6	(2.5)	109.0	(4.0)
All India	100.0		100.0		100.0		100.0		100.0		100.0	
CV	0.18		0.18		0.18		0.20		0.19		0.16	

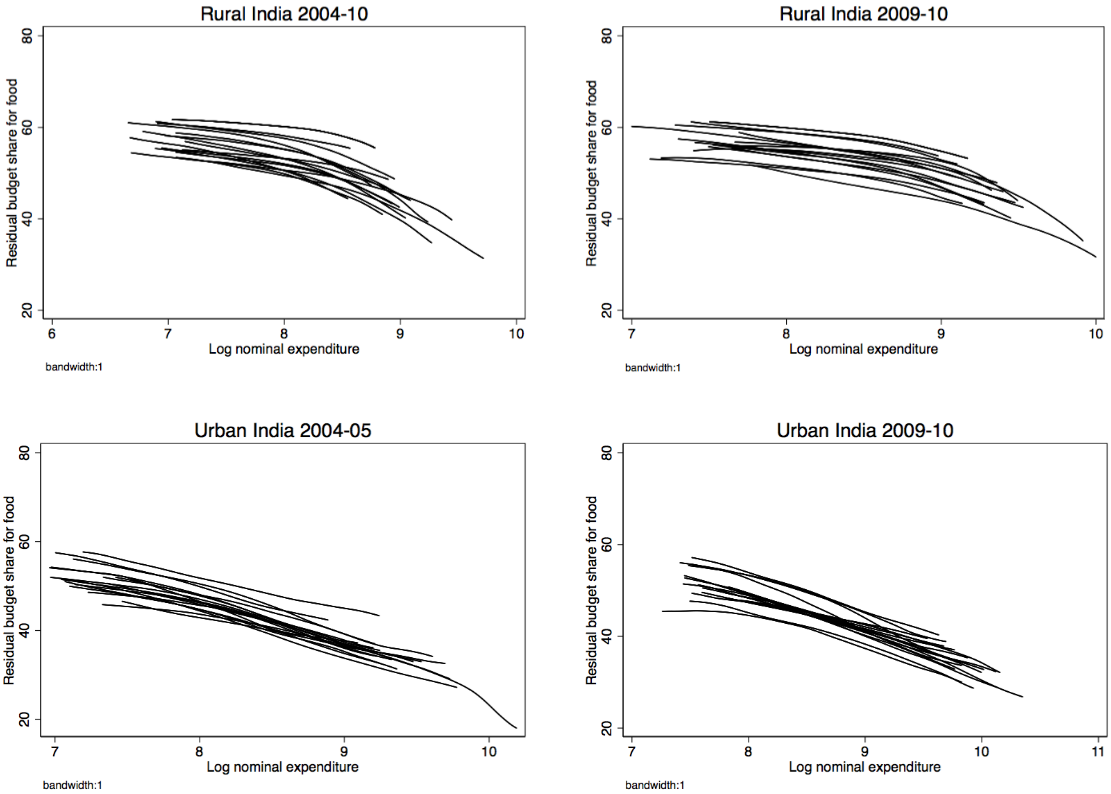
Note: The standard errors for the nonlinear price expressions depend, among other things, on the covariance between the b-coefficient and the dummy coefficients. This covariance is stronger when we use the instrument variable approach, which is the main reason for why the standard errors in Column (12) are smaller than those in Column (2).

Table A.6: All-India intertemporal prices

	Rural		**Urban**	
	(1)	(2)	(3)	(4)
Engel	161.4	(2.7)	160.6	(2.5)
Engel_{rp}	161.5	(2.9)	158.1	(2.7)
Engel_{sbs}	161.3	(3.0)	161.9	(2.6)
QEngel	159.1	(2.9)	160.7	(2.5)
Engel_{fs}	154.6	(1.3)	159.2	(1.4)
Engel_{fs-eq}	155.1	(1.3)	159.3	(1.4)
Engel_{IV}	161.1	(3.3)	161.2	(2.2)
UV (IPC)	150.6		150.3	
CPI*	155.0		145.1	

Note: *sbs* denotes the state-by-state analysis. Standard errors in parenthesis. * The CPIAL and CPIIW are used for the rural and urban sectors, respectively.

Figure A.1: Semi-parametric analysis of the Engel-relation



Note: The figures display estimates from the Epanechnikov kernel smoother, obtained using a bandwidth of unity and based on data on households comprising two children and two adults from the 17 major states. We remove the effect of all the covariates, except the logarithm of nominal expenditure, using the tenth-order optimal differencing weights proposed by Yatchew (2003). For the purpose of presentation, the figures are constructed after excluding the top and bottom one per cent of the expenditure distribution in each state and sector (urban/rural).

Appendix B Calorie consumption

Table B.1 displays All-India prices per 1000 calorie for some aggregated food groups. We derive these measures by computing a weighted average over the items in each group, using average budget shares as weights.

Figure B.1 shows the proportion of all calories consumed obtained from cereals (black lines) and edible oils & fats (grey lines), separately for urban and rural sector and each time period, for all states used in the calorie analysis.

Figure B.2 presents graphs based on a semi-parametric analysis which further explores this idea of similar calorie shares among the poor. In this case, we first remove any effects on the cereal-calorie shares from a set of covariates which may influence calorie patterns for equally poor households. This is done by differencing. The covariates include household demographics, occupation, the number of meals taken outside the home (for which we do not observe the calorie content) and an indicator for whether the household gets more than 50 per cent of its calories through the Public Distribution System (PDS), a government program for distributing subsidized food grains. We also construct an index to capture regional variation in prices per calorie for cereals relative to other food items.[31] We then plot the residual cereal-calorie shares against the logarithm of nominal expenditures. As before, we find no systematic differences across households close to our estimated poverty lines while this is not true of those close to either of the official poverty lines.

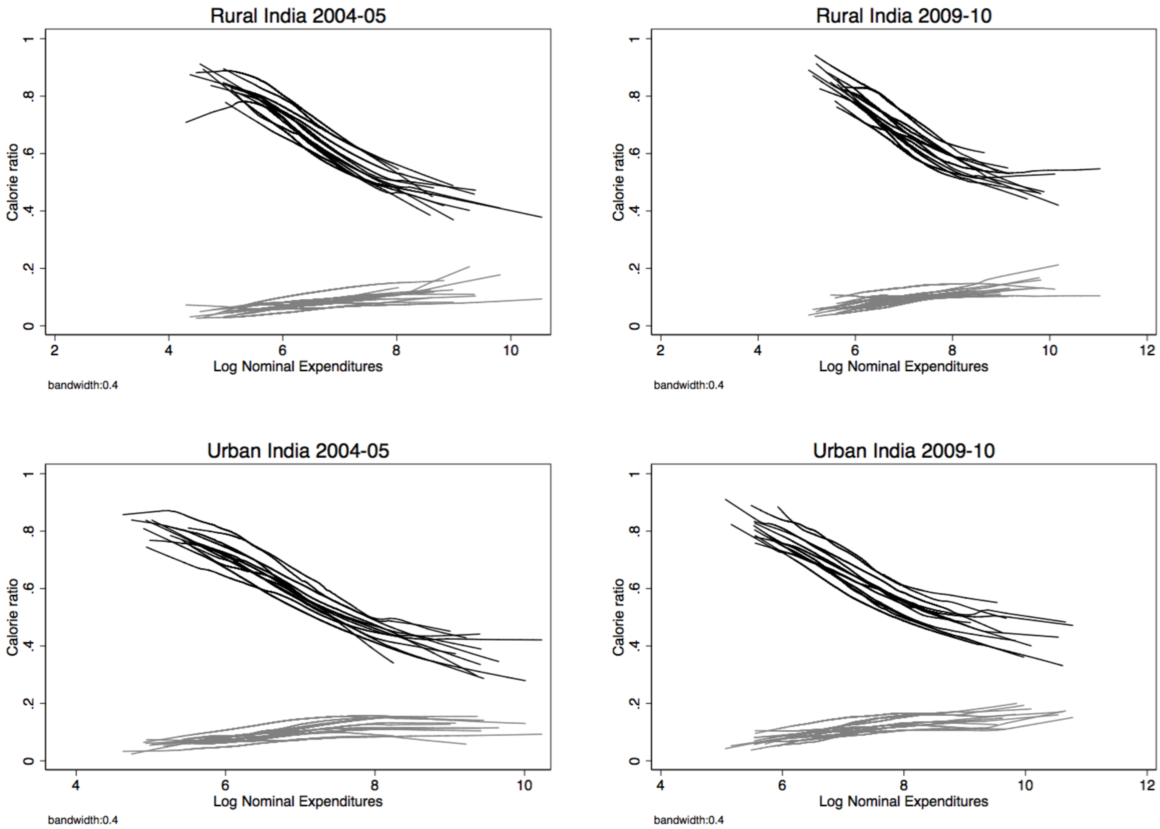
The bottom panel of the figure displays the outcomes from a similar exercise using the oil-calorie shares.

Table B.1: Prices per 1000 calorie (rupees)

	Rural		**Urban**	
	2004-05	2009-10	2004-05	2009-10
	(1)	(2)	(3)	(4)
Cereals and substitutes	3.1	5.0	3.7	6.4
Roots and tubers	12.8	24.0	15.0	26.7
Sugar and honey	9.5	15.4	12.5	18.2
Pulses and nuts and oilseeds	9.5	19.5	14.1	22.3
Vegetables and fruits	34.7	64.1	43.2	76.5
Meat, eggs and fish	58.8	104.9	68.4	111.9
Milk and milk products	21.2	31.4	24.0	32.4
Oils and fats	8.5	11.7	8.3	11.4
Misc. food, food products and beverages	39.2	58.0	40.0	61.0

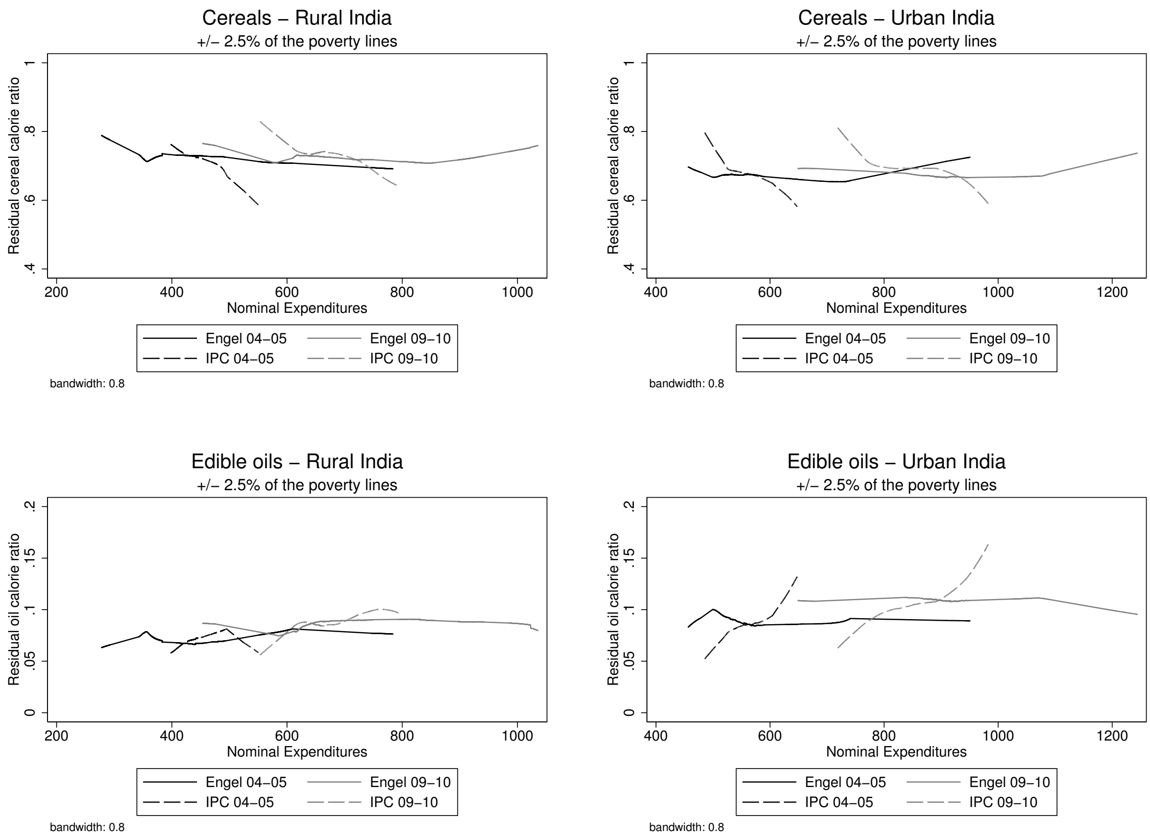
[31] We do this using the weighted country-product-dummy method (WCPD) due to Rao (1990). This is the same procedure as we use to construct the relative food/non-food price index in Section 5 and is explained in more detail in Appendix C.

Figure B.1: Calorie shares and the log of nominal expenditure



Note: The figures display estimates from the Epanechnikov kernel smoother, using a bandwidth of 0.4.

Figure B.2: Calorie shares and nominal expenditure levels



Note: The graphs in the figure display estimates from the semi-parametric analysis. The effect of the covariates is removed using the tenth-order optimal differencing weights proposed by Yatchew (2003).

Appendix C Relative food and non-food prices

We calculate unit values at the household level by dividing expenditure by quantity for each consumption item. The 61st NSS survey round provides information on quantities and values for 187 goods. Eleven of these items are found to be of insignificant value and are excluded in the Planning Commission methodology.[32] We exclude the same eleven items. To derive a comparable set of unit value items for the two survey rounds, we make additional adjustments. Items that appear in the questionnaires from the 66th round but not in those from the 61st are either excluded or aggregated with a relevant item. Items without readily available quantities in either of the two rounds, or items with non-comparable unit measures, are excluded.[33] Furthermore, following the official methodology, we aggregate PDS items with the relevant non-PDS items before calculating unit values.[34]

Given the set of household estimates we compute median unit values within each NSS district, separately for rural and urban areas and for each survey round. We then proceed by aggregating these medians into a food and a non-food price index using the weighted country-product-dummy method (WCPD) of Rao (1990). As there is no guarantee that every item is consumed in every district we choose this method that allows, and fills in for, missing observations. We obtain the aggregated indices by running the following weighted regression for food items and non-food items separately:

$$ln\ (median\ uv_{i,dst}) = \sum_i b_i D_i + \sum_d \sum_s \sum_t \alpha_{dst} D_{dst}, \tag{9}$$

where $ln\ median\ uv_{i,dst}$ is the median unit value for each good i, D_i is a dummy variable for every item i, and D_{dst} is a set of dummy variables for each district d, state s and time period t. As weights in the regression we use the item-wise average budget shares.[35] Finally, the aggregate food and non-food price estimates for each district are found directly from the dummy coefficients as:

$$P_{dst} = e^{\alpha_{dst}}. \tag{10}$$

This gives us two district-specific price indices. We derive the relative food/non-food index by dividing the food index by the non-food index. Finally, we normalize these relative price indices

[32] These are khoi, barley, singara, berries, misri, ice, katha, snuff, cheroot, ganja and cotton.

[33] Ice cream, other milk products and other intoxicants are excluded because of missing quantities in NSS66. Dhoti and sari are excluded because of non-comparable units (meters in NSS61 and numbers in NSS66). Soya beans are excluded from NSS61 because of their exclusion from NSS66. Petrol and diesel are excluded from NSS66 because of their exclusion from NSS61. Supari and lime are aggregated into other ingredients for pan in NSS61 because of their exclusion from NSS66. Second-hand footwear is aggregated into other footwear, and cooked meals received as assistance or payment are aggregated into cooked meals purchased in NSS66.

[34] This applies for rice, wheat, sugar and kerosene.

[35] We normalize these budget shares such that the sum of the covered items equals unity.

such that the population weighted all-India value equals unity in both rural and urban sectors in 2004–05.

Summary statistics are shown in Table C.1. Due to space considerations, we only report the average values for the major states. The table suggests that there are relatively large differences in relative food/non-food prices across Indian states. It can also be seen that the food unit values generally increased by more than the non-food unit values did during the five-year period from 2004–05 to 2009–10. The relative price index increased by roughly 15 per cent in the rural sector, and by 12 per cent in the urban sector. For comparison, the corresponding ratios increased by 21 per cent and 17 per cent in the CPIAL (rural) and CPIIW (urban) indices, respectively.[36]

TABLE C.1: Relative food and non-food prices

	Rural				**Urban**			
	2004–05		2009–10		2004–05		2009–10	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Andhra Pradesh	1.04	(0.12)	1.11	(0.12)	0.96	(0.11)	1.04	(0.11)
Assam	1.09	(0.11)	1.29	(0.10)	1.11	(0.13)	1.29	(0.14)
Bihar	0.89	(0.12)	1.07	(0.09)	0.90	(0.10)	1.01	(0.11)
Chhattisgarh	1.05	(0.06)	1.11	(0.20)	1.05	(0.08)	1.02	(0.14)
Gujarat	1.12	(0.12)	1.29	(0.15)	1.03	(0.10)	1.22	(0.12)
Haryana	0.99	(0.11)	1.17	(0.12)	0.94	(0.11)	1.11	(0.13)
Jharkhand	1.02	(0.10)	1.15	(0.09)	1.02	(0.15)	1.01	(0.10)
Karnataka	0.93	(0.07)	1.06	(0.13)	0.96	(0.07)	1.01	(0.12)
Kerala	1.07	(0.12)	1.15	(0.10)	1.01	(0.09)	1.06	(0.10)
Madhya Pradesh	1.01	(0.12)	1.12	(0.17)	1.04	(0.13)	1.04	(0.13)
Maharashtra	1.11	(0.15)	1.23	(0.18)	1.07	(0.12)	1.19	(0.12)
Odisha	0.98	(0.10)	1.23	(0.20)	0.97	(0.09)	1.12	(0.10)
Punjab	0.83	(0.10)	1.03	(0.16)	0.84	(0.09)	1.01	(0.16)
Rajasthan	0.95	(0.15)	1.24	(0.14)	0.93	(0.12)	1.12	(0.12)
Tamil Nadu	1.17	(0.11)	1.21	(0.10)	1.13	(0.09)	1.24	(0.12)
Uttar Pradesh	0.96	(0.10)	1.13	(0.11)	0.96	(0.13)	1.16	(0.15)
West Bengal	0.96	(0.07)	1.11	(0.10)	0.95	(0.05)	1.15	(0.08)
All India	1.00	(0.14)	1.15	(0.15)	1.00	(0.13)	1.12	(0.15)

Note: All values given are population-weighted state averages, obtained using the multipliers from the NSS data. The weighted all-India average for 2004–05 is normalized to unity for the rural and urban sectors separately. The standard deviations clustered at the district level are shown in brackets.

[36] These figures are obtained by comparing the food component with the non-food component in the two price indices.

Appendix D Group averages as instruments

Let F^* denote true food expenditures and Y^* denote total expenditures, and $y^* = logY^*$. The economic model is:[37]

$$\left(\frac{F^*}{Y^*}\right)_i = a + by_i^* + dD_i + \varepsilon_i^*, \tag{11}$$

where ε^* is a structural shock or omitted variable which is independent of y^*.

A potential problem is that both food expenditure and total expenditure may be measured with noise. Let us assume that the measurement errors are multiplicative so that we observe: $F_i = e^{\mu_i^F} F_i^*$ and $Y_i = e^{\mu_i^Y} Y_i^*$, where μ^x is some measurement error on x (which is independent of x^* and $E(x) = 0$).

Let's then consider the relationship estimated by 2-SLS: $\left(\frac{F}{Y}\right)_i = a + by_i + dD_i + \varepsilon_i$, where we use Z, the group mean value of y, as an instrument for y. Then the following moment conditions will hold:

$$\frac{1}{N}\sum_i (Z_i - \overline{Z})\left(\left(\frac{F}{Y}\right)_i - \hat{a} - \hat{b}y_i - \hat{d}D_i\right) = 0 \tag{12}$$

$$\frac{1}{N}\sum_i (D_i - \overline{D})\left(\left(\frac{F}{Y}\right)_i - \hat{a} - \hat{b}y_i - \hat{d}D_i\right) = 0 \tag{13}$$

$$\frac{1}{N}\sum_i \left(\left(\frac{F}{Y}\right)_i - \hat{a} - \hat{b}y_i - \hat{d}D_i\right) = 0 \tag{14}$$

We can solve the last condition for $\hat{a}$ and insert into the two other moment conditions, which gives:

$$\hat{b} = \frac{\sum_i (Z_i - \overline{Z})((\frac{F}{Y})_i - \overline{(\frac{F}{Y})}) \sum_i (D_i - \overline{D})(D_i - \overline{D}) - \sum_i (D_i - \overline{D})((\frac{F}{Y})_i - \overline{(\frac{F}{Y})}) \sum_i (D_i - \overline{D})(Z_i - \overline{Z})}{\sum_i (Z_i - \overline{Z})(y_i - \overline{y}) \sum_i (D_i - \overline{D})(D_i - \overline{D}) - \sum_i (D_i - \overline{D})(y_i - \overline{y}) \sum_i (D_i - \overline{D})(Z_i - \overline{Z})} \tag{15}$$

$$\hat{d} = \frac{\sum_i (D_i - \overline{D})((\frac{F}{Y})_i - \overline{(\frac{F}{Y})}) \sum_i (y_i - \overline{y})(Z_i - \overline{Z}) - \sum_i (D_i - \overline{D})(y_i - \overline{y}) \sum_i ((\frac{F}{Y})_i - \overline{(\frac{F}{Y})})(Z_i - \overline{Z})}{\sum_i (Z_i - \overline{Z})(y_i - \overline{y}) \sum_i (D_i - \overline{D})(D_i - \overline{D}) - \sum_i (D_i - \overline{D})(y_i - \overline{y}) \sum_i (D_i - \overline{D})(Z_i - \overline{Z})}. \tag{16}$$

To investigate whether our prices are consistent when instrumenting, we need to calculate the

[37] For illustrational purposes we only consider one dummy variable and no other explanatory variables.

probability limit of the price expression:

$$plim_{N\to+\inf}P = e^{-plim_{N\to+\inf}\left(\frac{\hat{d}}{\hat{b}}\right)} = e^{-\frac{cov(F/Y,D)cov(y,Z)-cov(F/Y,Z)cov(y,D)}{cov(F/Y,Z)var(D)-cov(F/Y,D)cov(Z,D)}}$$
$$= e^{-\frac{cov(e^{\mu F-\mu Y}F^*/Y^*,D)cov(y,Z)-cov(e^{\mu F-\mu Y}F^*/Y^*,Z)cov(y,D)}{cov(e^{\mu F-\mu Y}F^*/Y^*,Z)var(D)-cov(e^{\mu F-\mu Y}F^*/Y^*,D)cov(Z,D)}}$$
$$= e^{-\frac{cov(D,e^{\mu^F-\mu^Y}(a+by^*+dD+\varepsilon^*))cov(y,Z)-cov(Z,e^{\mu^F-\mu^Y}(a+by^*+dD+\varepsilon^*))cov(y,D)}{cov(Z,e^{\mu^F-\mu^Y}(a+by^*+dD+\varepsilon^*))var(D)-cov(D,e^{\mu^F-\mu^Y}(a+by^*+dD+\varepsilon^*))cov(Z,D)}}. \tag{17}$$

Now, assuming that μ^F, μ^Y and ε^* are independent of Z and D, this boils down to:

$$plim_{N\to+\inf}P = e^{-\frac{bE(e^{\mu F-\mu Y})(cov(y,Z)cov(y^*,D)-cov(y^*,Z)cov(y,D))+dE(e^{\mu F-\mu Y})(var(D)cov(y,Z)-cov(Z,D)cov(y,D))}{bE(e^{\mu F-\mu Y})(var(D)cov(y^*,Z)-cov(Z,D)cov(y^*,D))}}$$
$$= e^{-\frac{dE(e^{\mu F-\mu Y})(var(D)cov(y^*+\mu^Y,Z)-cov(Z,D)cov(y^*+\mu^Y,D))}{bE(e^{\mu F-\mu Y})(var(D)cov(y^*,Z)-cov(Z,D)cov(y^*,D))}} = e^{-\frac{d}{b}}, \tag{18}$$

which is equal to the true price level in our model.