CDE
January 2015

A BEHAVIOUR-BASED APPROACH TO THE ESTIMATION OF POVERTY IN INDIA

Ingvild Almas
Email: ingvild.almas@nhh.no
Norwegian School of Economics
and Business Administration
and University of Oslo, Bergen

Anders Kjelsrud
Email: a.g.kjelsrud@econ.uio.no
ESOP, University of Oslo

Rohini Somanathan
Email:rohini@econdse.org
Department of Economics
Delhi School of Economics

Working Paper No. 226
(REVISED)

Centre for Development Economics
Department of Economics, Delhi School of Economics

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

Ingvild Almås, Anders Kjelsrud and Rohini Somanathan†

Abstract

Almost a sixth of the world's population and a large fraction of its poorest people live in India. Until recently, national poverty estimates were widely criticized because they relied on aggregate price indices. A new official methodology based on micro price data was adopted in 2011. We propose an alternative approach based on the notion that comparably poor households can be identified through the proportion of their incomes spent on food. Compared with official estimates, we find higher levels of poverty in eastern India, and smaller reductions in poverty from 2005 to 2010. Our estimates have weaker data requirements than official methods. They also compare favorably on several validation tests: Households around our state poverty lines exhibit similar patterns of calorie consumption, rates of self-reported hunger are higher in states we classify as poor and we find a higher correlation of price levels across areas where arbitrage is likely. (*JEL*: D1, E31, F01, I32)

**Acknowledgements*: We thank Orazio Attanasio, Ragnhild Balsvik, Richard Blundell, Gernot Doppelhofer, Abhiroop Mukhopadhyay, Ragnar Nymoen and Fabien Postel-Vinay for useful comments. This paper is part of the research activities at the ESOP centre at the Department of Economics, University of Oslo. ESOP is supported by The Research Council of Norway.

†Almås: Norwegian School of Economics and University of Oslo, Bergen, email: `ingvild.almas@nhh.no`; Kjelsrud: ESOP, University of Oslo, email: `a.g.kjelsrud@econ.uio.no`; Somanathan: Dehli School of Economics, email: `rohini@econdse.org`.

1 Introduction

Almost a sixth of the world's population and a large fraction of its poorest people live in India. Indian poverty estimates are clearly crucial inputs in understanding world poverty trends, yet there is much disagreement about the legitimacy of methods used to derive poverty lines. When the official methodology for estimating poverty in India changed in 2011, the revised poverty lines resulted in a 50 per cent increase in rural head counts. The poverty debate has been particularly charged because of the mismatch between the rhetoric of poverty eradication and performance in spite of considerable economic growth and also because government programmes are targeted at families that are officially classified as poor.

Poverty lines can differ either because of alternative definitions of adequacy or variations in the cost of the subsistence bundle. The Indian controversy has been mainly about costs, although questions of adequacy have recently entered the discourse. Poverty lines were linked to calorie norms in the late 1970s and were subsequently adjusted using aggregate price indices for each of the Indian states, and separately for rural and urban areas. Over the years, this methodology was widely criticized: The rural-to-urban price differentials implicit in the lines were considered too large to be credible, the consumption weights used in the price indices were only infrequently updated, and the poverty lines failed to preserve the original calorie norms.

Although poverty measurement in most countries relies on some aggregation of local prices, there are well-documented deficiencies in the methods used for constructing price indices and in the quality of underlying price data (Costa, 2001; Deaton, 2010; Diewert, 1978; Hamilton, 2001; Neary, 2004; Nuxoll, 1994).[1] The methods recently adopted by the Indian government were aimed at overcoming many of these anomalies, but they are also now challenged. These methods use Fisher indices which are geometric averages of the so-called Laspeyres and Paasche indices. Neither the Laspeyres nor the Paasche index allows for substitution between goods in the consumption basket, a limitation that causes both indices to be affected by the so-called substitution bias. Taking the geometric average of the two limits the substitution bias, but does not eliminate it. Another feature of the

[1] See Chen et al (2008) for an overview over methods used in various countries.

new methods is that in the absence of micro price data, unit values are used to construct cost-of-living indices. As unit values are expenditure divided by quantity, they may reflect prices, but also quality of the products. Also, unit values only cover a subset of all goods, mainly foods, for which well-defined quantity units exist in the data. Because of these limitations, the debate on poverty measurement remains open and the search for a clear and defensible methodology for poverty measurement continues.

In this paper, we suggest an alternative method to estimate poverty that circumvents direct micro price measurement and aggregation of such micro price measures. We identify cost-of-living through the estimation of Engel curves for food. 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. Assuming that households with the same demographic and occupational characteristics spend the same proportion of their income on food, we use differences in nominal expenditures of households with the same food share to estimate relative price levels. We then use those relative price levels to derive poverty lines for the rural and urban parts of each state in each of the above years. We exploit the Engel relationship using Indian survey data to determine cost-of-living differences across states and across two time periods: 2005 and 2010.

Several studies have used estimates of Engel curves to correct for biases in the measurement of prices over time. Hamilton (2001) pioneered this strand of literature through his study on consumer price indices in the United States. 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); Olivia and Gibson (2012) have all applied this method to other countries and years. More recently, it has also been used to estimate biases in spatial price variations (Almås, 2012). 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 propose a method to do so (we later refer to this as the modified Engel method).

Our data come from National Sample Surveys (NSS), the standard source for poverty studies in India. Although consumer expenditure surveys are conducted each year, the large surveys that can be used for state-level estimates are typically quinquennial, hence the choice of 2005 and 2010 for our analysis.[2] These data are richer than those used in many of the studies cited above, and allow us to control for demographic and occupational variables that are likely to influence the budget share for food. This makes our identifying assumption of a stable Engel curve more plausible. Our price estimates are identified directly from the estimated Engel curves, and corresponding poverty measures are derived. We normalize our price estimates to yield the same aggregate price level for 2005 as official methods use to allow for a meaningful comparison of the poverty rates.[3]

We present three main results on poverty trends and patterns. First, the dispersion in price levels and corresponding poverty rates across Indian states exceeds that of official estimates. Second, poverty rates in the rural part of the eastern states of Assam, Bihar, Odisha and West Bengal are consistently higher than those implied by official figures and exceed 50 per cent in both survey years. Third, the decrease in overall poverty over our five-year period is much more modest than suggested by official statistics. All of these results appear robust to alternative empirical specifications of the Engel relationship and to changes in the composition of our sample.

The Engel method is a structural approach that rests on quite strong identifying assumptions. Although the structural assumptions are founded upon consumer choice theory and seem consistent with observed behavior in micro data, the method has been criticized, particularly for its failure to tackle certain forms of household heterogeneity.[4] First, even if the structural assumption about consumer behavior, i.e., the way the budget share for food changes with income, is consistent with micro data, it may still be the case that any two households may achieve different welfare levels, even if they have identical budget share for food and real expenditure level. This critique of the use of demand system estimation to infer welfare was raised by Pollak and Wales in several papers where they focused on the use of traditional equivalence scale adjustments to address heterogeneity in household composition (Pollak and Wales, 1979, 1981). As the number of children is partly a choice variable, and because households may differ in the welfare valuation of children, we cannot compare the welfare level of a household with three children to the

[2] As an exception to this practice, an extra survey was conducted in 2011 to update poverty measures.

[3] Chattopadhyay (2010) and Coondoo *et al.* (2011) are also related in that they use NSS consumption data to estimate prices. Their approach is however more structural, estimating a larger set of parameters and relying on grouped rather than individual micro data.

[4] Note that standard price indices can be criticized along the same lines.

welfare level of a household with two children simply by using a measure of real income and some chosen equivalence scale. In this paper, we have chosen to focus on households with two children and two adults in order to avoid problems related to using a traditional equivalence scale adjustment. But, because we are worried about the potential selection effect of choosing to study households of a specific composition and size, we also provide a series of robustness checks where we include all households using controls for children and adults and different equivalence scales, to calculate poverty. All the checks confirm our main results on cost of living measures and poverty calculation. However, the results of these robustness checks should be interpreted with caution, keeping the critique raised by Pollak and Wales in mind.

Second, there may be residual heterogeneity even among households of the same composition and size. In particular, heterogeneity arising from the fact that different households live in different geographical areas has received some attention in the literature. A recent paper by David Atkin discusses habit formation leading to differences in preferences across geographical areas in India. Atkin (2013) does not study preference differences across food and non-food - the categories used for identification in this paper - but rather within the food consumption category. Yet his main argument that there *may* be unobserved heterogeneity and differences in preferences that *may* impact the results from the Engel method is still valid. Third, as we know that preferences are non-homothetic, if relative prices differ, and if relative prices affect the budget share for food, cost of living is not only different across geographical areas, but also between households with different income levels. As the Engel curve only provides one cost of living for each geographical area (in each time period), it can only identify the cost of living of one particular income level. As we are concerned with identifying the cost of living for those around the poverty line this in itself is not a problem. However, as shown in Beatty and Crossley (2012), the income level for which the Engel method identifies cost of living, is unknown to the researcher and hence it is not guaranteed that the cost of living we report is that for the income level around the poverty line. All these questions call for a validation of the results from the Engel curve method.

In addition to this paper's findings about poverty, a contribution is to provide a validation of the results from the Engel method. The data we use allow us to conduct an empirical investigation of the Engel based cost of living and poverty estimates as well as the new and old official poverty counts and implicit cost of living indices. First, we examine correlations between estimated prices in urban and rural areas within states. Markets are likely to be well integrated within states leading us to expect a positive correlation between these two sets of prices. Deaton and Tarozzi (2000) compute such correlations based on unit values from the same survey data for previous periods and find them to be high and positive.[5] The correlation coefficients of our estimated rural and urban prices are positive for both our survey years and higher than those resulting from the current official methodology for estimating prices and poverty.

Our second and most elaborate validation check examines the sources of calories for households with per capita consumption in a narrow band around our estimated poverty lines. One might expect that the poor 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). Our survey data indicate that cereals are the cheapest sources of calories and that their share of calories falls as household expenditure increases. If our estimated poverty lines do represent the same level of real household income, we would expect households around these lines in different states to have similar cereal-calorie shares. When we limit our sample to households in a symmetric five per cent interval around the poverty lines for each state and for each of the two time periods, we find that households clustered around our estimated lines get similar shares of their total calories from cereals. This is not true for the official lines, which suggests that we have been able to more accurately identify equally poor households across states and time periods. As a final check we examine rates of self-reported hunger across states and find the highest rates in many of the states that we classify as the poorest.

Our paper also compares estimates from the current and previous official methodology for estimating poverty used by the Indian government and finds the new methods constitute a significant improvement on previous techniques. For example, we show that the new estimates of price levels within states exhibit much higher correlations than before, and the spatial distribution of poverty seems more consistent with those from other studies (see e.g. Deaton and Tarozzi, 2000; Deaton, 2003). As such we show that the new official

[5] Using data for the two NSS rounds, 1987-1988 and 1993-1994, they find this correlation to be close to .7 in both years.

estimates perform quite well in the validation exercises, far better than the old outdated official numbers. But, we also show that despite potential challenges to the identification through the Engel method, the method provides very well in the validation exercises, and substantially better than the current improved official methodology.

The rest of this paper is organized as follows. In Section 2, we outline a brief chronology of poverty measurement in India. In Section 3, we describe our empirical methodology in detail and discuss the data and the variables used in the main analysis. The key findings are presented in Section 4. In Section 5, we present results on cereal-calorie shares and the validation checks of our approach. In Section 6, we report results from a range of specification checks. Concluding remarks are provided in Section 7.

2 Indian poverty measures: a chronology

Poverty lines during the colonial period and in the decades immediately following independence were quite arbitrary and based on varying notions of adequacy (Srinivasan, 2007). In 1979 the notion of subsistence was first 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).[6] 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.[7] In 1993, an expert group set up by the Planning Commission recommended state-specific poverty lines based on regional

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

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

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 and Tarozzi, 2000; Deaton, 2003, 2008). Another expert committee was formed in late 2005 led by Suresh Tendulkar and most of its suggestions were 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 head count for 2004–05 to 25.7 per cent, the official estimate under the old procedure. Using this normalization, 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, but excludes education, health care and other services. For these categories of consumption, price information was obtained from a variety of sources.[8] This makes the new procedures somewhat ad hoc and difficult to replicate in the future.[9]

Based on the new methods, rural poverty was found to be 50% higher than previous estimates for 2004-05 and regional patterns of poverty were also quite different. In 2011, the same methodology was used to compute poverty estimates based on the NSS survey

[8] The cost of school attendance is derived from the NSS employment–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.

[9] Subramanian (2011) provides a critical review of the new methodology.

of 2009-10. We therefore have two sets of official estimates for 2004-05, corresponding to the old and new methods adopted by the Indian Planning Commission for estimating poverty. For 2009-10, we have estimates based only on the new methods suggested by the Tendulkar Committee. We compare our results to both sets of official numbers in Section 4.

The new official estimates continued to be controversial and led the Indian Planning Commission to constitute yet another expert group to evaluate them in June 2012. The group's recommendations were released in June 2014. It is in this context that we believe a more theoretically grounded approach to measurement is especially valuable. We present this in the next section.

3 Data and Methods

Challenges related to the quality of local price data and to techniques used for its aggregation have spurred a literature that uses structural methods to estimate price differences (see for example, Neary, 2004). Two strands of this literature are especially relevant for us. Hamilton (2001) first relied on estimated Engel curves for food over time to identify biases in the consumer price index and then used the corrected prices to deflate nominal income for the United States. This method has since been widely applied to other countries and time periods.[10] Almås (2012) used the same Engel relationship in a spatial setting to identify biases in purchasing power parity numbers of the Penn World Tables.

The method we propose here is a variation on the above approaches. Instead of using shifts in Engel curves to identify biases in existing prices as has previously been done, we directly estimate relative prices through the systematic variation in Engel curves for similar households over the two years and for each Indian state. 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.

[10]The most significant of these studies have been referred to in the Introduction.

We estimate the following demand system (Deaton and Muellbauer, 1980a):[11]

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

where m_{hst} is the budget share for food, 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.[12]

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:

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

where 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 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.[13]

An attractive feature of this approach is that we can identify price variation without re-

[11] We assume here that the budget share for food is not influenced by relative prices. This is discussed later in this section and relaxed in Section 6.

[12] Note that if we aimed at identifying 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 Pendakur, 2002; Almås and Sørensen, 2012).

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

lying directly on price data. Our main specification assumes that the budget share for food is not influenced by relative prices, and we are able to identify comparable price levels by using expenditure data and demographics, only. We relax this assumption in Section 6 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.[14] The 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 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).

We use data from two recent large NSS rounds 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.[15] Summary statistics, covering 222,558 households, are shown in Table 1. Consumption expenditures are recorded based on a 30-day recall period for most consumption goods and on some infrequently purchased items using a 365-day recall period.[16] The NSS values items received in-kind at their average

[14]The evidence on the effect of relative prices on food shares in mixed. 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.

[15]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.

[16]This applies to clothing, bedding, footwear, education, institutional medical expenses and durable goods. 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.

local retail price, while home production is evaluated at market prices net of transport costs.

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 on the urban and rural samples separately.

Table 1: Household descriptive statistics

	Rural				Urban			
	2004–05		2009–10		2004–05		2009–10	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Monthly per capita expenditure (Rupees)	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.

4 Results

Table 2 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).[17]

Table 2: Demand system estimates

Dep. var.: Budget share for food (%)	**Rural**	**Urban**
Log of household expenditure	-12.628	-13.761
	(0.215)	(0.181)
Observations	14257	9112
R^2	0.379	0.522

Note: Robust standard errors 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 cultivated land categories listed in Table 1.

We use the coefficients for log expenditures and the state-year dummies to calculate prices based on Equation 4. Although our estimates are based on the full sample of 30 states and union territories described above, we present results for the 17 largest states labelled as *major states* by the NSS. These cover roughly 80 per cent of the Indian population.[18] To obtain the prices implicit in the official poverty lines, we divide the official poverty lines by the all-India poverty line for each sector and time period. These spatial price indices are in columns (1)-(8) of Table 3 under the headings *Engel* and IPC_c, the c subscript indicating that the estimates are based on the current official methods as opposed to those initially used to compute the 2004-05 poverty lines (standard errors for the Engel based price estimates are reported in Table 7 and 8). 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 in the last row, there is more price variation in rural as compared to urban areas under both methods and the Engel

[17] The expenditure elasticities are calculated as $1 + \frac{b}{m}$ where m is the mean food share in the sample.

[18] We focus on these 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.

prices imply more dispersion than official measures in both sectors.

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, an average annual increase of approximately 9 per cent.[19]

Given these price indices, it is straightforward to compute updated poverty lines and head counts. 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.[20] We then derive state poverty lines for both time periods by multiplying the all-India lines for 2004–05 with the state prices in Columns (1), (3), (9) and (11) divided by 100. This procedure implies that our estimated all-India head-count ratios for 2004–05 differ from the official ones only because of different spatial prices while the head counts for 2009–10 deviate on both spatial and intertemporal dimensions.

Table 4 presents head counts based on the Engel analysis together with those from current and 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.

[19]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 9 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.

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

TABLE 3: 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_c	*Engel*	IPC_c	*Engel*	IPC_c	*Engel*	IPC_c	*Engel*	IPC_c	*Engel*	IPC_c
	(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: CV denotes the coefficient of variation. The all-India values are population-weighted averages of state-level prices, normalized to 100. The subscript c denotes that we are using the current official poverty measures.

Table 4: Poverty head counts

	2004–05						2009–10			
	Rural			**Urban**			**Rural**		**Urban**	
	Engel	IPC_c	IPC_p	*Engel*	IPC_c	IPC_p	*Engel*	IPC_c	*Engel*	IPC_c
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Andhra Pradesh	32.0	32.3	10.5	20.2	23.4	27.4	27.7	22.7	16.0	17.7
Assam	85.8	36.2	22.1	45.4	21.8	3.6	73.7	39.9	43.6	25.9
Bihar	82.3	55.7	42.7	59.9	43.7	36.1	76.9	55.3	64.9	39.4
Chhattisgarh	29.0	55.1	40.8	35.4	28.4	42.2	17.3	56.1	12.5	23.6
Gujarat	37.4	39.1	18.9	19.9	20.1	13.3	46.5	26.6	21.6	17.7
Haryana	11.6	24.8	13.2	15.2	22.4	14.5	26.4	18.6	28.2	23.0
Jharkhand	78.1	51.8	46.3	36.7	23.8	20.3	40.1	41.3	47.1	31.0
Karnataka	23.4	37.5	20.7	18.5	25.9	32.6	34.5	26.1	16.9	19.5
Kerala	8.3	20.2	13.2	23.1	18.4	20.0	3.5	12.0	18.6	12.1
Madhya Pradesh	17.4	53.6	36.8	25.7	35.1	42.7	19.0	42.0	16.3	22.8
Maharashtra	20.3	47.9	29.6	15.9	25.6	32.1	20.2	29.5	15.4	18.3
Odisha	63.0	60.8	46.9	46.2	37.6	44.7	53.0	39.2	41.3	25.9
Punjab	4.6	22.1	9.0	9.0	18.7	6.3	3.1	14.6	17.5	18.1
Rajasthan	30.1	35.8	18.3	28.7	29.7	32.3	33.9	26.4	26.0	19.9
Tamil Nadu	46.3	37.5	23.0	17.4	19.7	22.5	20.5	21.2	18.2	12.7
Uttar Pradesh	27.9	42.7	33.3	37.8	34.1	30.1	32.3	39.3	39.7	31.7
West Bengal	66.8	38.3	28.4	40.0	24.4	13.5	70.3	28.8	37.3	21.9
All India	39.7	41.8	28.3	25.6	25.7	25.7	37.7	33.3	24.8	20.9

Note: The all-India rates are weighted averages of the state-level poverty head counts, using the NSS population multipliers. The subscript c and p denote current and previous official poverty measures respectively.

5 Validating our estimates

We have seen that our method yields prices and corresponding poverty rates that differ–substantially for some states–from the official measures. Is there any reason to believe that our poverty numbers are more credible than officially published ones? In this section we present the outcome of three quite different exercises that lead us to have confidence in the validity of our estimates.

Our first exercise investigates the correlation between the rural and urban price indices derived from our estimates. 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 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 old Planning Commission measures which exhibit a *negative* correlation between spatial prices in rural and urban areas (-0.34 in 2004–05).[21] This seems implausible and suggests that the price measures in use until recently were out of date.[22]

Our second exercise is the most elaborate of the three and 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. If we believe that poor families are likely to maximize their calorie intakes, one would also expect them to rely on cheap sources of calories. With increases in income, they 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

[21] This correlation coefficient is based on the spatial prices implied by the previous official poverty lines. Note that these poverty lines were estimated for only 22 of the 30 states and union territories. The spatial price index is therefore based on these 22 states only.

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

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.[23] In Appendix A (Table 10), 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.

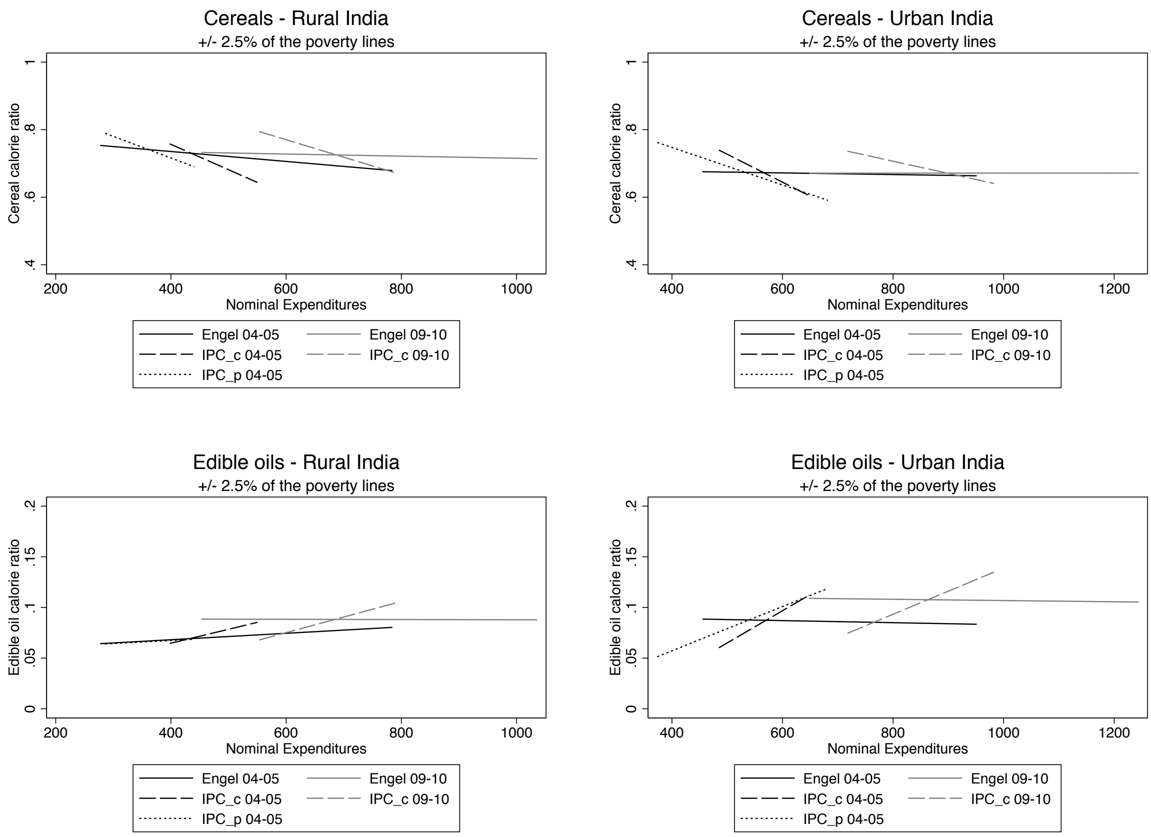
We use this negative relationship between cereal-calorie shares and income to evaluate the Engel-based and official poverty counts. We do this by examining the cereal-calorie shares of households in a symmetric 5 per cent interval around the two sets of poverty lines. If the state-wise poverty lines 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.[24]

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, based on the current and the previous methodology. 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

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

[24]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.

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.

to act as if they were poorer than households close to the poverty lines in other states.[25] This suggests that the official methods fail to capture real cost-of-living differences across Indian states. In Appendix A we conduct a semi-parametric analysis, which indicates that these findings are not driven by differences in relative food prices or other observed household characteristics. 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.[26]

[25] 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.

[26] 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).

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 a potential candidate. In Appendix A we show that there is a positive relationship between edible oil-calorie shares and the logarithm of total expenditure, although this relationship is much 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 the official lines, oil-calorie shares are rising in nominal income.[27]

The above analysis has used the set of poverty lines derived in Section 4. 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 5 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% value corresponds to the slopes in Figure 1.

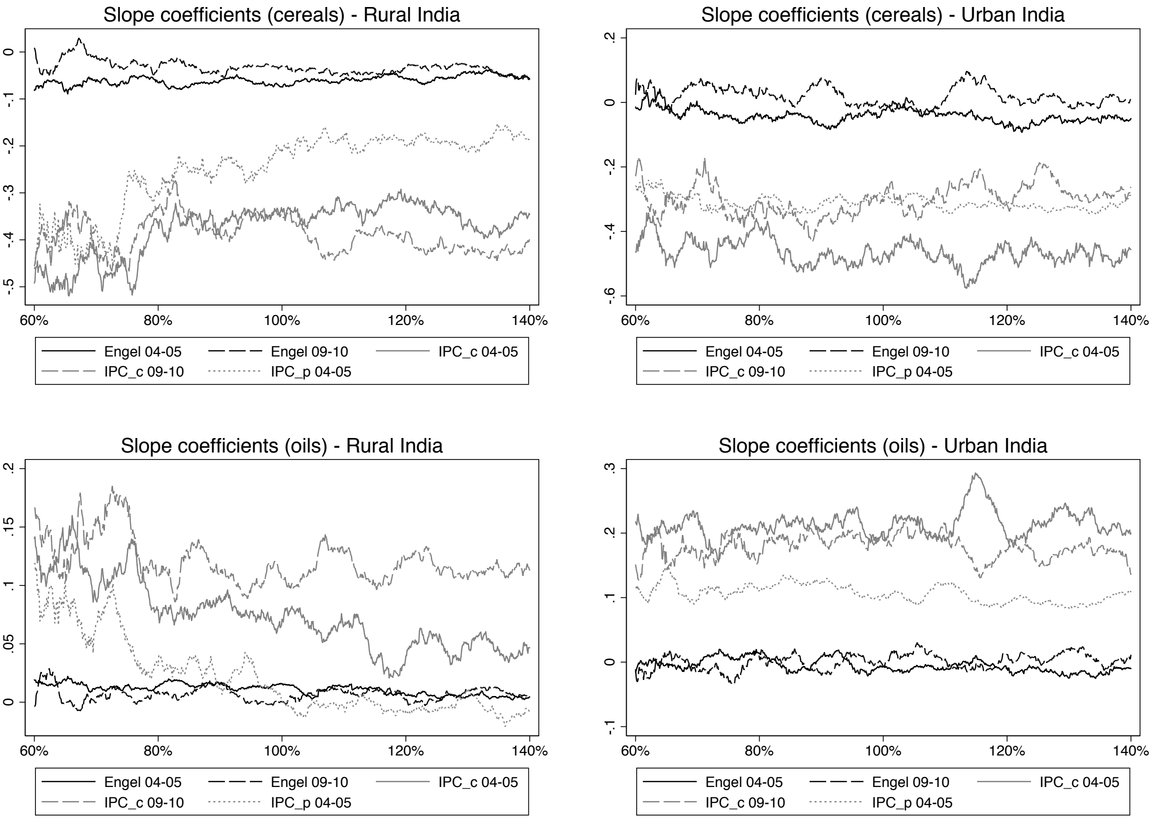
For all scalar multiples of the poverty lines we use, the slopes are roughly zero for the Engel lines for both cereals and oils. This is reassuring both for our estimates of the current pattern of poverty but also as validations of this procedure for future poverty lines, which may rely on a definition 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 minor and final validity check of our estimates, we use responses on household perceptions of hunger. In the NSS survey, respondents are asked whether every member of the household gets "enough food every day".[28] This is a self-reported measure of

[27]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.

[28]These proportions are taken from the "type 2" NSS survey, because the question does not appear in

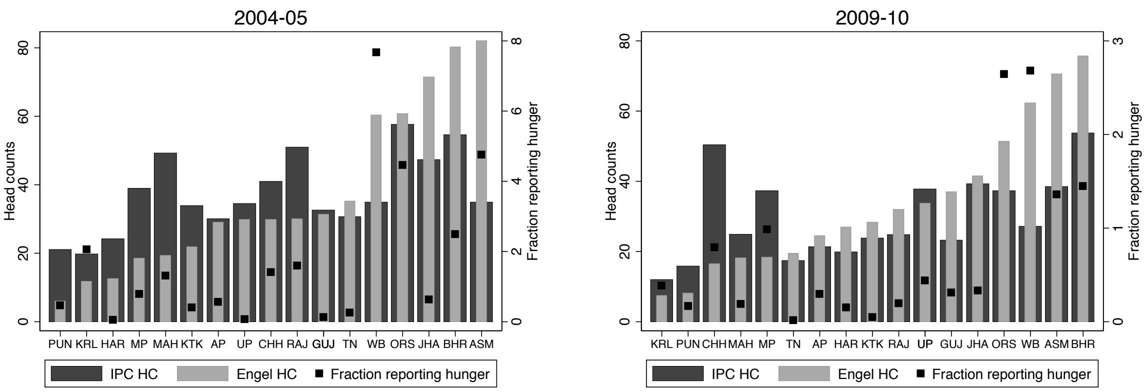
FIGURE 2: Slope coefficients calorie analysis for different poverty lines



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

hunger and should be interpreted with the usual caveats. However, we have little reason to expect any systematic errors across states.[29] Figure 3 shows the proportion of all households reporting a lack of food.[30] These numbers are plotted against two sets of head-count ratios: those from the Engel analysis and the new 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 Engel-based poverty estimates are higher than official numbers.

Figure 3: Head counts 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 5 shows the overall correlation between head-counts and self-reported hunger ratios. The poverty counts based on the new official methodology are positively correlated with the self-reported hunger but the correlations are smaller than those for the Engel counts. For urban areas the negative correlation between hunger and poverty counts based on the old official methodology further suggests that the previous Planning Commission measures are misleading and out-dated.

the "type 1" survey that we use for the rest of our data.

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

[30] We combine the rural and urban head counts, using population weights.

TABLE 5: Correlations between self-reported hunger and head counts

Correlation coefficients	**Engel**		**IPC$_c$**		**IPC$_p$**	
	Rural	Urban	Rural	Urban	Rural	Urban
2004–05	0.55	0.74	0.20	0.28	0.30	-0.15
2009–10	0.64	0.53	0.41	0.41		

6 Robustness analysis

In this section, we perform several checks on our specification and our sample. We first include 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.[31] This ratio is constructed from

[31]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

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.[32] 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 in Appendix B.

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

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

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. We see from Table 7 and 8 that these estimates are almost identical to our earlier price estimates.

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 sector to unity we can pick up the price level in the second period by estimating:

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

and using the dummy-coefficient, d to compute:

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

The fourth row in Table 9 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

reason to expect that including relative prices and an alternative reference utility level is quantitatively important.

[32]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.

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 4, 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):

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

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 fourth column of Table 7 and 8 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 fifth row of Table 9 reports the implied inter-temporal price measures. These are very similar to those from our main specification.

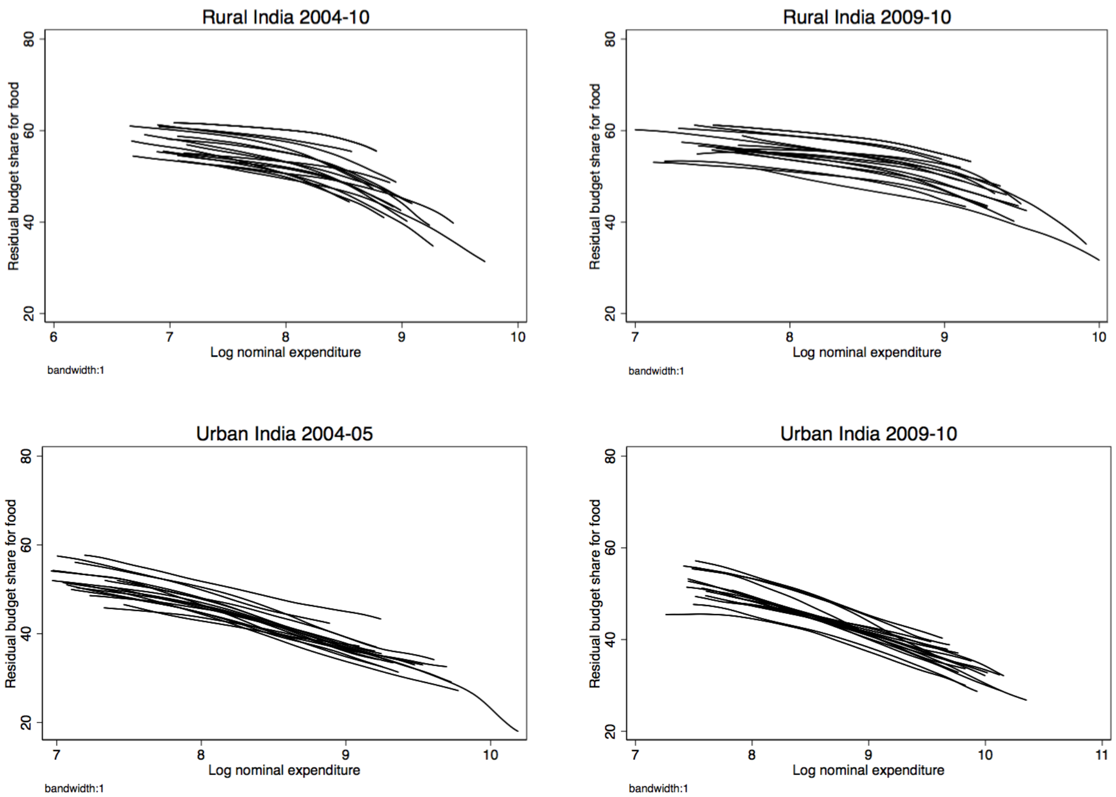
In Section 4, we restricted our sample to households with two children and two adults. Estimates from the full sample of households generate similar expenditure elasticities for the urban sector and slightly larger elasticities for the rural sector.[33] The corresponding spatial prices are presented in Column (3) of Table 7 and 8. 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.

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

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.[34] 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 C for details). As shown in the last column in Table 7 and 8, this gives very similar spatial patterns as in the main specification. We also see from Table 9 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.

FIGURE 4: 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.

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

TABLE 6: Regressions for robustness checks

	Rural				Urban			
Dep. var.: Budget share for food (%)	$Engel_{rp}$	$Engel_{fs}$	$QEngel$	$Engel_{IV}$	$Engel_{rp}$	$Engel_{fs}$	$QEngel$	$Engel_{IV}$
Log of household expenditure	-12.628	-11.165	38.113	-10.229	-13.761	-13.102	6.168	-15.498
	(0.215)	(0.0852)	(2.524)	(0.418)	(0.181)	(0.079)	(3.000)	(0.386)
Log of household expenditure squared			-3.219				-1.222	
			(0.161)				(0.183)	
Log of relative food/non-food prices	-0.085				1.902			
	(0.623)				(0.786)			
Observations	14257	137152	14257	14257	9112	85300	9112	9112
R^2	0.379	0.348	0.395	0.371	0.523	0.472	0.524	0.516

Note: Robust standard errors are given in parentheses. 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 7: Robustness checks: Spatial prices rural

	$Engel$ (1)		$Engel_{rp}$ (2)		$Engel_{fs}$ (3)		$Engel_{fs-eq}$ (4)		$QEngel$ (5)		$Engel_{iv}$
2004–05											
Andhra Pradesh	96.8	(2.3)	96.9	(2.3)	99.6	(1.2)	92.2	(2.4)	93.9	(2.9)	
Assam	171.2	(5.4)	171.4	(5.4)	183.4	(2.2)	186.7	(9.6)	192.4	(8.5)	
Bihar	126.5	(3.5)	126.3	(3.6)	129.5	(1.3)	133.5	(6.0)	137.8	(5.2)	
Chhattisgarh	71.0	(3.0)	71.0	(3.0)	76.4	(1.3)	68.1	(3.3)	68.4	(3.7)	
Gujarat	109.6	(4.0)	109.7	(4.1)	111.8	(1.7)	104.3	(4.5)	105.7	(4.9)	
Haryana	97.9	(6.2)	97.9	(6.2)	98.3	(2.2)	99.3	(5.0)	88.9	(7.2)	
Jharkhand	117.0	(4.8)	117.0	(4.8)	116.9	(1.8)	132.8	(9.5)	128.5	(6.7)	
Karnataka	82.7	(2.8)	82.7	(2.9)	84.3	(1.1)	78.8	(3.0)	78.6	(3.5)	
Kerala	93.5	(3.8)	93.5	(3.9)	89.4	(1.5)	94.2	(3.1)	82.0	(4.7)	
Madhya Pradesh	63.8	(2.5)	63.8	(2.5)	59.8	(0.8)	62.7	(2.1)	59.2	(3.0)	
Maharashtra	78.6	(2.4)	78.7	(2.4)	71.6	(0.8)	76.6	(2.2)	72.5	(3.0)	
Orissa	93.3	(3.5)	93.3	(3.5)	101.6	(1.4)	98.1	(4.5)	98.6	(4.5)	
Punjab	88.8	(4.2)	88.7	(4.3)	86.7	(1.6)	89.0	(3.9)	78.2	(5.1)	
Rajasthan	100.6	(4.2)	100.6	(4.2)	98.3	(1.4)	97.9	(3.8)	98.4	(5.1)	
Tamil Nadu	107.6	(3.4)	107.7	(3.5)	100.6	(1.4)	105.9	(3.5)	107.4	(4.2)	
Uttar Pradesh	83.9	(2.7)	83.8	(2.7)	81.3	(0.7)	80.8	(2.3)	80.4	(3.3)	
West Bengal	133.1	(3.4)	133.1	(3.4)	139.1	(1.4)	135.5	(4.8)	142.1	(4.8)	
All India	100.0		100.0		100.0		100.0		100		
CV	0.22		0.22		0.26		0.24		0.27		
2009–10											
Andhra Pradesh	102.6	(2.9)	102.6	(2.9)	105.0	(1.4)	101.2	(2.8)	100.2	(3.5)	
Assam	141.8	(5.6)	141.9	(5.7)	165.8	(2.5)	152.7	(8.2)	157.4	(8.2)	
Bihar	117.1	(4.0)	117.0	(4.0)	122.1	(1.4)	122.4	(5.4)	125.4	(5.5)	
Chhattisgarh	62.6	(4.4)	62.6	(4.4)	60.5	(1.4)	65.3	(3.4)	57.8	(5.0)	
Gujarat	122.5	(5.3)	122.6	(5.3)	117.1	(2.0)	116.1	(5.9)	123.2	(6.7)	
Haryana	123.4	(7.9)	123.5	(7.9)	119.0	(2.8)	124.1	(7.0)	119.2	(9.6)	
Jharkhand	85.1	(4.2)	85.1	(4.2)	103.1	(1.9)	84.4	(4.5)	86.6	(5.3)	
Karnataka	95.3	(3.7)	95.3	(3.7)	86.2	(1.4)	93.3	(4.1)	92.4	(4.5)	
Kerala	82.0	(3.7)	82.0	(3.7)	75.8	(1.4)	91.8	(3.2)	68.9	(4.6)	
Madhya Pradesh	65.2	(2.8)	65.2	(2.8)	65.1	(1.0)	66.6	(2.5)	60.3	(3.3)	
Maharashtra	92.6	(2.8)	92.6	(2.8)	81.7	(1.0)	89.8	(2.7)	87.7	(3.5)	
Orissa	91.5	(3.4)	91.5	(3.4)	95.3	(1.5)	94.2	(3.9)	94.9	(4.4)	
Punjab	91.9	(5.5)	91.8	(5.5)	95.7	(2.2)	90.2	(4.5)	82.3	(6.4)	
Rajasthan	114.4	(5.5)	114.5	(5.5)	103.5	(1.6)	111.2	(5.1)	115.6	(6.8)	
Tamil Nadu	88.7	(2.8)	88.8	(2.8)	82.7	(1.3)	85.5	(2.5)	83.6	(3.5)	
Uttar Pradesh	87.8	(2.9)	87.8	(2.9)	90.3	(0.9)	85.0	(2.6)	86.5	(3.6)	
West Bengal	129.0	(3.7)	128.9	(3.7)	131.4	(1.7)	131.0	(5.1)	139.0	(5.2)	
All India	100.0		100.0		100.0		100.0		100		
CV	0.21		0.21		0.24		0.22		0.28		

Note: The subscript rp denotes relative price controls, fs denotes that the full sample is used in the estimation, and iv denotes that the instrument variable approach is used. Robust standard errors in parenthesis. Standard errors for the nonlinear price expressions are computed using the delta method.

Table 8: Robustness checks: Spatial prices urban

	$Engel$ (1)		$Engel_{rp}$ (2)		$Engel_{fs}$ (3)		$QEngel$ (4)		$Engel_{iv}$ (5)	
2004–05										
Andhra Pradesh	91.1	(2.5)	91.7	(2.5)	89.5	(1.3)	90.0	(2.5)	91.0	(2.2)
Assam	160.5	(8.2)	158.5	(8.1)	150.1	(3.4)	162.7	(9.7)	151.2	(7.1)
Bihar	120.0	(6.7)	121.8	(6.8)	127.9	(2.3)	121.7	(7.4)	113.6	(5.8)
Chhattisgarh	96.8	(5.8)	96.2	(5.7)	84.5	(2.0)	96.5	(6.5)	95.6	(5.2)
Gujarat	113.3	(3.8)	112.7	(3.8)	115.9	(1.6)	112.2	(4.4)	113.6	(3.4)
Haryana	94.0	(4.7)	94.5	(4.7)	92.6	(1.8)	95.0	(5.0)	96.9	(4.4)
Jharkhand	117.5	(7.0)	117.8	(7.0)	132.1	(2.7)	121.0	(7.5)	114.0	(6.2)
Karnataka	89.0	(2.7)	89.5	(2.7)	87.4	(1.3)	87.9	(3.1)	90.7	(2.4)
Kerala	111.1	(5.1)	110.5	(5.1)	101.0	(1.9)	109.2	(4.5)	113.3	(4.6)
Madhya Pradesh	80.7	(2.9)	80.3	(2.9)	77.3	(1.1)	80.3	(3.1)	81.4	(2.6)
Maharashtra	88.4	(2.0)	87.6	(2.0)	90.3	(0.8)	88.8	(2.1)	90.8	(1.9)
Orissa	99.8	(4.8)	100.3	(4.8)	112.5	(2.5)	100.5	(5.0)	96.6	(4.2)
Punjab	92.7	(3.6)	95.1	(3.8)	90.8	(1.4)	93.4	(3.9)	94.9	(3.3)
Rajasthan	96.7	(4.6)	97.7	(4.6)	100.1	(1.6)	97.1	(4.6)	97.6	(4.2)
Tamil Nadu	92.1	(2.3)	90.4	(2.4)	98.0	(1.2)	91.5	(2.2)	92.3	(2.0)
Uttar Pradesh	98.0	(3.3)	98.5	(3.3)	95.4	(1.1)	98.8	(3.3)	97.1	(2.9)
West Bengal	125.0	(4.2)	125.6	(4.3)	121.5	(1.6)	124.2	(4.7)	120.6	(3.7)
All India	100.0		100.0		100.0		100.0		100	
CV	0.20		0.21		0.22		0.21		0.19	
2009–10										
Andhra Pradesh	97.0	(2.9)	98.0	(3.0)	100.3	(1.3)	97.1	(2.6)	98.4	(2.7)
Assam	132.0	(7.9)	129.1	(7.8)	129.0	(3.1)	133.4	(8.3)	125.2	(6.8)
Bihar	115.7	(5.8)	117.3	(5.9)	110.8	(2.0)	117.8	(6.4)	109.9	(5.0)
Chhattisgarh	67.5	(5.2)	68.1	(5.3)	69.7	(1.9)	70.2	(4.5)	70.2	(5.0)
Gujarat	110.3	(4.4)	108.8	(4.4)	110.9	(1.7)	108.8	(4.7)	110.2	(3.9)
Haryana	114.3	(5.3)	114.6	(5.3)	107.7	(2.1)	114.7	(5.4)	115.4	(4.8)
Jharkhand	113.6	(6.5)	114.3	(6.6)	117.0	(2.7)	116.1	(7.0)	109.7	(5.7)
Karnataka	93.2	(3.5)	94.2	(3.5)	96.7	(1.4)	91.9	(3.4)	93.6	(3.1)
Kerala	100.8	(5.6)	101.5	(5.7)	92.3	(1.8)	100.7	(4.3)	103.6	(5.2)
Madhya Pradesh	72.2	(2.9)	73.0	(3.0)	68.9	(1.0)	72.5	(2.8)	74.2	(2.7)
Maharashtra	98.1	(2.9)	97.6	(2.8)	99.8	(1.0)	98.3	(2.6)	99.7	(2.6)
Orissa	99.4	(6.1)	99.3	(6.1)	101.0	(2.3)	99.4	(5.4)	96.8	(5.3)
Punjab	102.4	(4.1)	104.2	(4.3)	103.1	(1.8)	101.9	(4.6)	103.1	(3.7)
Rajasthan	99.9	(4.8)	99.8	(4.8)	104.5	(1.8)	99.7	(4.7)	100.2	(4.3)
Tamil Nadu	96.3	(2.5)	95.1	(2.6)	94.0	(1.2)	95.9	(2.5)	96.6	(2.3)
Uttar Pradesh	97.4	(3.5)	97.1	(3.5)	95.1	(1.1)	97.1	(3.3)	96.5	(3.1)
West Bengal	113.6	(4.2)	113.6	(4.2)	111.2	(1.4)	114.5	(4.4)	109.0	(3.7)
All India	100.0		100.0		100.0		100.0		100	
CV	0.18		0.18		0.19		0.18		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 (5) are smaller than those in Column (1).

TABLE 9: All-India intertemporal prices

	Rural		**Urban**	
Engel				
Engel	161.4	(2.4)	160.6	(2.2)
Engel_{rp}	161.5	(2.7)	158.1	(2.4)
Engel_{fs}	155.6	(0.8)	159.3	(0.9)
Engel_{sbs}	161.3	(2.7)	161.9	(2.4)
QEngel	159.1	(2.4)	160.7	(2.2)
Engel_{iv}	161.1	(3.0)	161.2	(2.0)
UV (IPC)	150.6		150.3	
CPI*	155.0		145.1	

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

7 Conclusion

In this paper, we have proposed a method for poverty comparisons in which price levels and poverty lines are estimated based on the behavioral assumption that equally poor households spend the same proportion of their incomes on food. This approach, based on the estimation of Engel curves for food, has been used in several contexts to correct for biases in prices. We apply it for the first time in estimating both spatial and inter-temporal variation in prices and corresponding poverty counts.

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 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 one set is more credible than the other. The methods we use to examine the validity of our estimates are an important methodological and empirical contribution of our paper. Our most important 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 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 estimates. Given the other advantages of our approach, namely its theoretical under-pinnings and relatively weak demands on data, we feel confident that it offers new possibilities for the better measurement of poverty and provides a benchmark against which to evaluate alternative methods.

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 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?

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*. Number 2006-2275. 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*, pages 418–430.

Deaton, A. (2003). Prices and poverty in India, 1987-2000. *Economic and Political Weekly*, pages 362–368.

Deaton, A. (2008). Price trends in India and their implications for measuring poverty. *Economic & Political Weekly*, pages 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 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. *Princeton, July*.

Dickens, R., Fry, V., and Pashardes, P. (1993). Non-linearities and equivalence scales. *The Economic Journal*, pages 359–368.

Diewert, E. (1978). Superlative index. *Econometrica: Journal of the Econometric Society*, **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.

Jensen, R. and Miller, N. (2010). A revealed preference approach to measuring hunger and undernutrition. Technical report, National Bureau of Economic Research.

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: Journal of the Econometric Society*, pages 694–703.

Neary, J. P. (2004). Rationalizing the Penn World Table: true multilateral indices for international comparisons of real income. *American Economic Review*, pages 1411–1428.

Nuxoll, D. A. (1994). Differences in relative prices and international differences in growth rates. *The American Economic Review*, **84**(5), 1423–1436.

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**.

Pollak, R. A. and Wales, T. J. (1979). Welfare comparisons and equivalence scales. *The American Economic Review*, pages 216–221.

Pollak, R. A. and Wales, T. J. (1981). Demographic variables in demand analysis. *Econometrica: Journal of the Econometric Society*, **49**, 1533–1551.

Rao, P. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association*, pages 163–171.

Srinivasan, T. (2007). Poverty lines in India: Reflections after the Patna conference. *Economic and Political Weekly*, pages 4155–4165.

Subramanian, S. (2011). The poverty line: Getting it wrong again. *Economic & Political Weekly*, **46**(48), 37.

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 Univ Pr.

Appendix A Calorie consumption

Table 10 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 5 shows the proportion of all calories consumed obtained from cereals (black lines) and edible oils & fats (grey lines), separately for each sector and time period, for all states used in the calorie analysis in Section 5.

Figure 6 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.[35] 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.

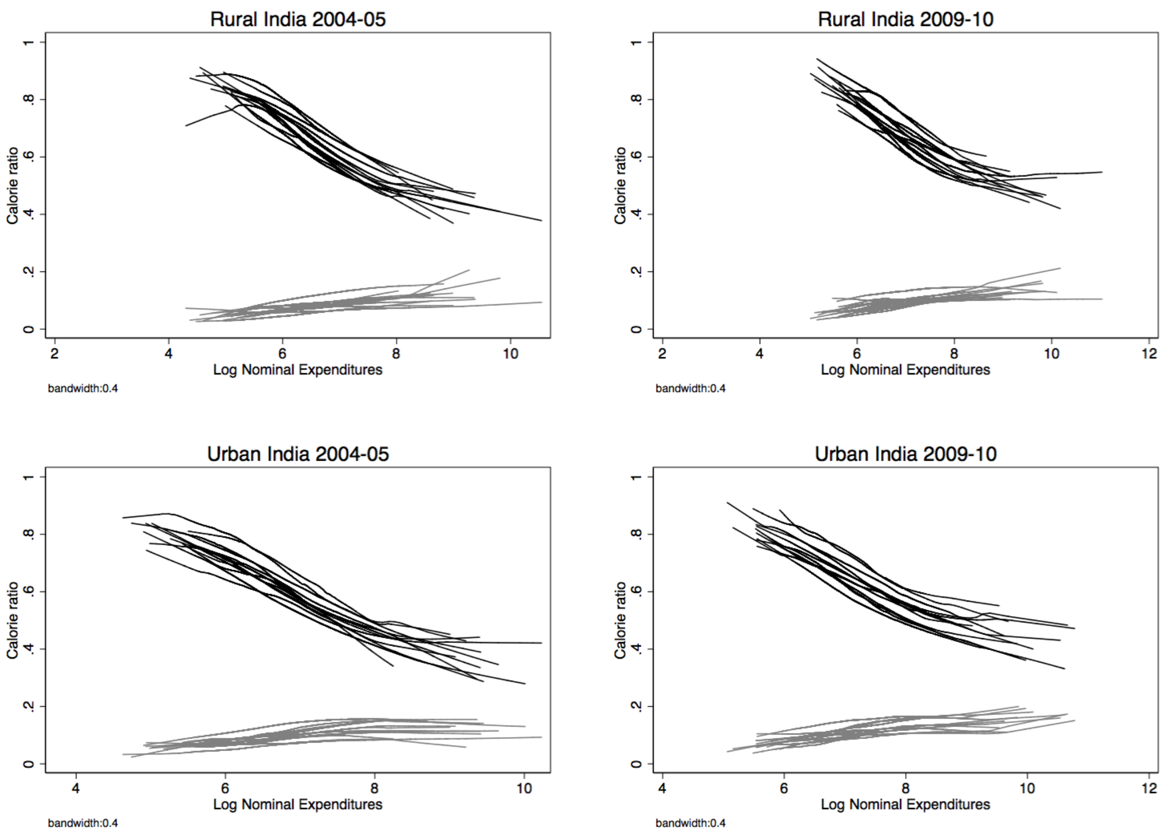
The bottom panel of the figure displays the outcomes from a similar exercise using the oil-calorie shares.

[35] 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 6 and is explained in more detail in Appendix B.

Table 10: Prices per 1000 calorie (rupees)

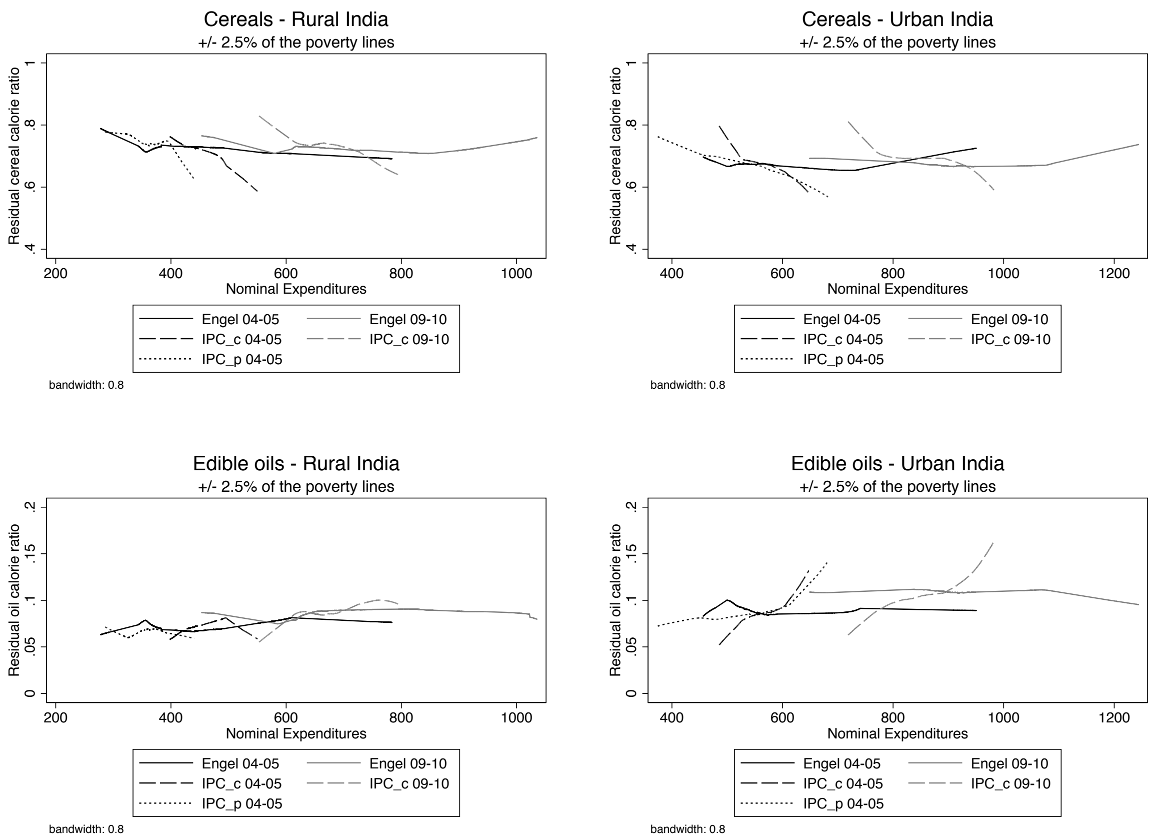
	Rural		Rural	
	2004-05	2009-10	2004-05	2009-10
Cereals and substitutes	3.1	5.0	3.7	6.4
Roots and tubers	12.8	24.0	15.0	26.7
Sugar and honey	9.5	15.4	12.5	18.2
Pulses and nuts and oilseeds	9.5	19.5	14.1	22.3
Vegetables and fruits	34.7	64.1	43.2	76.5
Meat, eggs and fish	58.8	104.9	68.4	111.9
Milk and milk 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

Figure 5: 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 6: 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 B 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.[36] 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.[37] Furthermore, following the official methodology, we aggregate PDS items with the relevant non-PDS items before calculating unit values.[38]

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 follow 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.[39]

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

[37] 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.

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

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

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 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 11. 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.[40]

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

Table 11: Relative food and non-food prices

	Rural				Urban			
	2004–05		2009–10		2004–05		2009–10	
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.

Appendix C Group averages (not for publication)

Let F^* denote true food expenditures and Y^* denote total expenditures, and $y^* = logY^*$. The economic model is:[41]

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

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: $\left(\frac{F}{Y}\right)_i = a + by_i + dD_i + \epsilon_i$, 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

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

the probability limit of the price expression:

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

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

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

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