Working Papers

www.cesifo.org/wp

Pro-poor Price Trends and Inequality – The Case of India

Ingvild Almås
Anders Kjelsrud

CESifo Working Paper No. 5740
Category 3: Social Protection
February 2016

ISSN 2364-1428

CESifo *Working Paper No. 5740*

Pro-poor Price Trends and Inequality – The Case of India

Abstract

It is well known that people's consumption patterns change with income. Relative price changes therefore affect rich and poor consumers differently. Yet, the standard price indices are not income-specific and hence, the use of these mask these differences in cost-of-living. In this paper, we study consumption inequality in India, while fully allowing for non-homotheticity. Our analysis shows that the changes in relative prices in a large part of the period from 1993 to 2012 were pro-poor, in the sense that they favored the poor relative to the rich. As a result, we also find that the standard measures significantly overestimate the rise in real inequality. Moreover, we show that the allowance for non-homotheticity is quantitatively much more important in our application than the adjustment for substitution in consumption, despite the larger attention paid to the latter in the literature. We also illustrate how conventional measures exaggerate inter-temporal changes in inequality when there is segregation in consumption / production, by which we mean that people's consumption patterns are skewed towards goods intensively produced by people of their own group.

JEL-codes: C430.

Keywords: inequality, cost-of-living, measurement.

Ingvild Almås
IIES
Stockholm University
Stockholm / Sweden
ingvild.almas@iies.su.se

Anders Kjelsrud
Statistics Norway
Research Department
Oslo / Norway
gkj@ssb.no

We are grateful to Erik Sørensen for sharing some of his Python scripts with us. 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, University of Oslo. ESOP is supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 179552.

1 Introduction

It is well known that people's consumption patterns change with economic affluence, i.e. that preferences are non-homothetic. Relative price changes will hence affect income groups differently even if all face the same set of prices (Muellbauer, 1974). Yet, the conventional price indices are not group specific, and if used to deflate individual incomes, they therefore mask these possible differences in cost-of-living.[1] This is likely to be a problem of first-order importance when discussing distributions and inequality, but it might also be problematic for other types of analysis as it is not very transparent whose cost of living the standard indices represent (see e.g., Beatty and Crossley, 2016). For example, the standard consumer price index formulae would generate price indices that represent the consumption of relatively rich individuals and due to the aggregation technique used, the larger the inequality in the society, the richer the "representative" household.

In this paper, we study consumption inequality in India, and we construct group-specific cost-of-living in order to do so. The period under study is 1993-94 to 2011-12 and our analysis shows that the changes in relative prices in a large part of this period were pro-poor, meaning that they favored the poor relative to the rich. Some of this effect is driven by relatively low price increases for food grains during the 1990s. We also show that the pattern of relative price changes has a large impact on measured inequality. Standard measures suggest that inequality rose steeply during the period studied (Datt and Ravallion, 2009; Weisskopf, 2011; The World Bank, 2011).[2] However, as much as about one-third of the increase between 1993-94 and 2004-05 disappears when we apply our expenditure specific cost-of-living adjustment. This finding illustrates that it is crucial to account for non-homotheticity when measuring consumption inequality. For the years after 2004-05, we find that the relative price changes were pro-rich and that the standard measures therefore somewhat understate the rise in inequality.

The standard price indices have other biases beside those induced by relying on homothetic preferences. For example, the fixed basket approaches, such as the Laspeyres, the Paasche and the classical Geary method—the latter underlying the Penn World Table—fail to incorporate substitution, as the assumed consumer basket is held fixed in comparisons involving different relative price levels. A large part of the price index literature is about how to avoid this problem (Akmal and Dowrick, 2005; Diewert, 1978; Feenstra *et al.*, 2012; Neary, 2004). In our empirical investigation, we make an effort to disentangle the biases caused by not adjusting for substitution and the biases caused by implicitly relying on homothetic preferences. This is done by comparing our estimates, which incorporate both substitution and non-homotheticity, with inequality measures derived through the Geary-index, which does not allow for either of the two, and with measures

[1] See Feenstra and Timmer (2015) for an overview of standard price indices used for comparisons of income/consumption.

[2] There is much less evidence on the trend in income inequality. As one of the few exceptions, Banerjee and Piketty (2005) present trends in top incomes and wages for the period 1922-2000 using individual tax return data.

derived through an index that allows for substitution but that relies on homothetic preferences. This comparison suggests that substitution alone has a very limited quantitative importance in our application—the differences between our estimates and the traditional fixed basket approaches are driven almost entirely by the allowance for non-homotheticity.

The analysis is conducted on household data collected by the National Sample Survey Organisation (NSS). This is the standard source for household expenditure comparisons in India.[3] Using these survey data, we construct the group specific cost-of-living indices in three main steps. In the first step, we calculate unit values and use those as measures of item prices (Deaton, 2008; Deaton and Dupriez, 2011; Deaton and Tarozzi, 2000). In the second step, we characterize consumer preferences. This is necessary in order to account for non-homotheticity. It is also necessary in order to incorporate substitution in consumption. As a way of recovering preferences, we estimate the Quadratic Almost Ideal Demand System (Banks *et al.*, 1997), using 11 aggregate consumption groups and percentiles of the expenditure distributions within each state, sector (urban and rural) and time period as the unit of observation.[4] The system is identified through spatial and inter-temporal variation in prices and variation in household consumption and under the assumption of homogenous preferences. The assumption of homogeneous preferences is restrictive, but even with this assumption, we allow for more heterogeneity, namely the heterogeneity in cost of living across groups, than any standard analysis of inequality. Future research should aim at also addressing heterogenous tastes in the measurement of cost of living and inequality. In the third and final step, we make use of the estimated price and income responses to compute money metric utilities and use that to calculate expenditure specific cost-of-living. In this latter computation, we make use of a Geary-like reference price vector, as suggested by Neary (2004) for the homothetic case, and by Almås and Sørensen (2012) for the non-homothetic case.

The money metric utilities are, in turn, used to construct real inequality measures. To evaluate the robustness of our findings, we repeat the procedure for a series of alternative specifications. All these alternative specifications provide similar inequality trends as the main analysis, and all confirm that the allowance for non-homotheticity is quantitatively more important than the allowance for substitution.

Our expenditure specific cost-of-living indices capture the effects of relative prices on the demand side and highlight that relative price changes affect people differently. Our findings also illustrate how conventional measures of real inequality are biased, depending on the particular patterns of relative price changes—if the prices of luxury goods increase relative to the prices of necessities, they will typically overestimate real inequality, and vice versa if the opposite is the case. We are not the first to discuss this type of bias. Some papers have, for example, proposed solutions to how

[3] We use the survey rounds from 1993-94, 1999-00, 2004-05, 2009-10 and 2011-12.

[4] The number of goods groups that we use is similar to many applications, e.g., the number of goods corresponds to that of the Penn World Table basic headings. Our findings are robust to categorizing goods in different ways, we have tested several groupings and our findings hold up.

to weight individual cost-of-living to obtain one aggregated "social cost-of-living index" (Crossley and Pendakur, 2010; Muellbauer, 1976; Pollak, 1980, 1981). More recently, other papers have directly discussed how price changes within countries affect different income groups (Cravino and Levchenko, 2015; Faber, 2012; Handbury, 2013; Moretti, 2013). Mishra and Ray (2011), Nicholas *et al.* (2010) and Pendakur (2002) investigate real consumption inequality in India, Australia and Canada, respectively, correcting for cost-of-living differences by indices closely related to ours. Similarly to what is done in this paper, these authors calculate money metric utility using the cost function. However, the other standard indices are not derived in any of the papers and they do not make an attempt to adjust for cost-of-living differences across geographical areas. Hence, they cannot nail down how important the adjustment for non-homotheticity is in comparisons to other adjustments. One of the contributions of our paper is to calculate cost-of-living deflators across time and space using standard indices and thus separate the bias stemming from the assumption of homothethic preferences from other types of biases.

Price changes, say exogenous changes from the world market, will also almost always change the distribution of nominal income. If the price of a good rises relative to other prices, this is likely to lead to higher income and expenditure levels for people working in the industry producing that particular good. This income effect should be present in all measures of real consumption inequality, as income is likely to be reflected in expenditure data. In contrast, and as discussed above, the full implications on cost-of-living are usually ignored. The fact that standard measures incorporate only one out of the two likely effects of relative price changes is problematic and might give rise to a systematic mismeasurement of the *variance* of inequality over time. In our application, we find that the standard inequality measures fluctuate more over time as compared to our measures that account for non-homotheticity. This is especially the case for the rural sector. One plausible hypothesis consistent with this is segregation in consumption/production. By this we mean that the effects on income and cost-of-living are systematically related, and that people's consumption patterns are skewed towards goods intensively produced by people of their own group—materialized in, for example, the poor producing and consuming necessities, and the rich, to a larger extent, producing and consuming luxury goods.[5]

We provide some indicative evidence of the above skewness by comparing households engaged in crop production with other types of households. The rural crop producers are poorer than the average household and their labor incomes react positively to increases in crop prices. Consequently, we find that decreases in the relative price of cereals (at the state level) are associated with increases in the conventional inequality measures. However, crop producers also devote a larger-than-average share of their budget towards food grains, not because they produce these particular goods, but simply because they are poor. Their cost-of-living therefore increases relative to the cost-of-living of richer households when the prices of grains rise. When we allow

[5]See Lewis (1954) for an extreme version of such a segregation.

for non-homotheticity, and hence adjust for the systematic difference in consumption patterns, the association between cereal prices and inequality is substantially weaker and no longer statistically significant. This particular example illustrates a more general phenomenon, namely that conventional measures that rely on homothetic preferences overstate changes in real inequality that are induces by relative price changes when there are elements of segregation in consumption/production.

The rest of the paper is organized as follows. In Section 2 we describe the construction of the different cost-of-living indices used in the empirical investigation. In Section 3 we present the data and discuss the implementation of our methods. We present our main findings in Section 4. In Section 5 we discuss the robustness checks, whereas concluding remarks are given in Section 6.

2 Non-homothetic preferences and cost-of-living

This section gives an overview of the different cost-of-living indices used in the analysis. For brevity, we use the notation "unit" for a unique state in a specific time period and sector (urban or rural). Throughout, there are n commodities indexed $i = 1, \ldots, n$, and m units indexed $j = 1, \ldots, m$. For each unit, there is a price vector p^j and a corresponding per capita quantity vector q^j. The total quantity consumed in a unit is given by the vector Q^j. Per capita nominal consumption in unit j is given by $z_j = p^j {q^j}^T$.[6]

The Geary index, also known as the Geary–Khamis index, is based on the idea of evaluating quantities, not by actual prices, but by a vector of average prices, π. The real per capita consumption level of unit j, evaluated in this way, could be written as:

$$I_j^{cons} = \pi q^j, \tag{1}$$

and the corresponding cost-of-living index as:

$$P_j^{cons} = \frac{p^j q^j}{\pi q^j}. \tag{2}$$

So far, this is similar to any conventional consumer price index. Therefore, we label this index by "cons", for "consumption index". As actual quantities are evaluated at the reference prices, this index does not take into account substitution in consumption. That is, the index does not adjust for the fact that the consumers would have chosen a different consumption basket if faced with the reference prices instead of the actual prices in their unit. The failure of the standard indices, such as the Geary index, to account for substitution has spurred a literature on more structural cost-of-living indices, sometimes referred to as "the economic approach" to price index

[6]For simplicity, we will skip the notation for transposed in the rest of the text and use e.g., $p^j q^j$ instead of $p^j {q^j}^T$.

measurement (Akmal and Dowrick, 2005; Neary, 2004).[7] This approach requires the estimation of preferences and is based on evaluating money metric utilities, $m(\pi, p^j, z_j)$. The real consumption level of unit j in this system could be denoted by:

$$I_j^{exp\text{-}h} = m(\pi, p^j, z_j) = e(\pi, v(p^j, z_j)), \tag{3}$$

where $e(.)$ and $v(.)$ are the expenditure function and the indirect utility function, respectively (that are specified once preferences have been estimated, more on this later). The cost-of-living index of unit j could now be written as:

$$P_j^{exp\text{-}h} = \frac{e(p^j, v(p^j, z_j))}{e(\pi, v(p^j, z_j))}. \tag{4}$$

The system allows for substitution in consumption, but does not allow for non-homotheticity. For this reason, we use the labelling, "exp-h", for "expenditure homothetic", where the expenditure part refers to the computation through the expenditure function. If prices differ across goods, and if the consumption basket changes with real income, there is no unique cost-of-living for every individual within a unit. The cost-of-living will not only depend on prices, but also on income. Even if we are only interested in the average cost-of-living in each unit, indices of the form in (4) cannot be applied, since there is no representative consumer when preferences are non-homothetic.

To fully allow for non-homotheticity, we construct a final real consumption index, $I_j^{exp\text{-}nh}$ for unit j, as:

$$I_j^{exp\text{-}nh} = L_j^{-1} \sum_{l=1}^{L_j} e(\pi, v(p^j, z_{jl})), \tag{5}$$

where z_{jl} gives per capita nominal consumption for individual l in unit j. The equation sums the money metric utilities for all individuals, $l = 1, \ldots, L_j$, in each unit. We label this extension by "exp-nh", for "expenditure non-homothetic", as it fully allows for non-homothetic preferences. The disaggregated nature of this index allows us to compute every individual's real consumption level from $e(\pi, v(p^j, z_{jl}))$ or, equivalently, by adjusting their nominal consumption level using the income-specific cost-of-living index:

$$P_{jl}^{exp\text{-}nh} = \frac{e(p^j, v(p^j, z_{jl}))}{e(\pi, v(p^j, z_{jl}))}. \tag{6}$$

The implementation of the above expenditure indices requires a procedure to determine the reference price vector, and a characterisation of preferences. Below we discuss both of these in turn.

[7]See also Almås (2012), Costa (2001) and Hamilton (2001) for related approaches.

In our main set of calculations, we determine the reference prices for all three indices in a Geary-like fashion. The Geary approach implicitly identifies reference prices by requiring that total consumption of each good should have the same overall value whether evaluated at the reference prices or at each unit's own prices, divided by their estimated cost-of-living. For the consumption index, this could be stated as follows:

$$\sum_{j=1}^{m} \pi_i Q_{ij} = \sum_{j=1}^{m} \frac{p_{ij} Q_{ij}}{P_j^{cons}}, \quad \text{for all } i = 1, \ldots, n. \tag{7}$$

These n linear equations in π determine the n reference prices (up to a normalization). Neary (2004) suggested a procedure to calculate similar types of reference prices in money metric cost-of-living indices. The procedure calculates the reference price vector π as in the classical Geary calculation, but multiplies the reference prices with virtual instead of actual quantities. The virtual quantities are those that would have been consumed if the reference prices had been the actual prices. For this reason, we are able to account for substitution. By Shepard's lemma, these quantities could be identified through the Hicksian demand functions. Thus, for the expenditure homothetic index, we could determine the reference prices by the following equations:

$$\sum_{j=1}^{m} \pi_i H_i(\pi, u_j) = \sum_{j=1}^{m} \frac{p_{ij} Q_{ij}}{P_j^{exp\text{-}h}}, \quad \text{for all } i = 1, \ldots, n, \tag{8}$$

where $H_i(\pi, u_j)$ is the total amount of virtual quantities of item i that would have been consumed in unit j at prices π. To take account of within-unit distributions of expenditures, the corresponding equations for the expenditure non-homothetic index become (Almås and Sørensen, 2012):

$$\sum_{j=1}^{m} \pi_i \sum_{l=1}^{N_j} h_i(\pi, u_{jl}) = \sum_{j=1}^{m} p_{ij} \sum_{l=1}^{N_j} \frac{q_{ijl}}{P_{jl}^{exp\text{-}nh}}, \quad \text{for all } i = 1, \ldots, n. \tag{9}$$

These two sets of nonlinear equations determine the reference prices in the two expenditure based systems, just as the (linear) equations in (7) determine the reference prices for the Geary system. In the robustness section, we propose yet two alternative procedures to determine the reference prices. All our main results hold up when using these alternative procedures.

To recover the necessary preference parameters, we estimate the Quadratic Almost Ideal Demand System (QUAIDS) due to Banks *et al.* (1997). The QUAIDS system is consistent with utility maximization. The budget share equation for good i can be expressed in the following flexible form:

$$\omega_{ij} = \alpha_i + \sum_{h=1}^{n} \gamma_{ih} \ln p_{hj} + \beta_i \ln y_j + \frac{\lambda_i}{\beta(p^j)} (\ln y_j)^2, \tag{10}$$

where $\ln y_j = \ln z_j - \ln \alpha(p^j)$, z_j is nominal per capita expenditure, and $\alpha(p^j)$ and $\beta(p^j)$ are price indices that depend on the parameters.[8] Moreover, the log expenditure function in the QUAIDS could be expressed as:[9]

$$\ln e(p_j, u_j) = \ln \alpha(p^j) + \frac{u_j \beta(p^j)}{1 - u_j \lambda(p^j)}. \tag{11}$$

The next section describes the data and the computation of the above cost-of-living indices.

3 Data and implementation

3.1 Data and price estimates

Our analysis is based on the nationwide household surveys collected by the National Sample Survey Organization (NSS). The NSS conducts household expenditure surveys every year, but the large surveys which can be used for state-level analysis are typically quinquennial. We use the five most recent such survey rounds, conducted in 1993–94, 1999–00, 2004–05, 2009–10 and 2011-12.[10] We limit the analysis to the 17 states labelled as "major" by the NSS. These states account for almost the entire Indian population.[11]

The household surveys include information on consumption expenditure for a wide range of items. However, to ease the estimation of the demand system, we aggregate all consumption items into 11 groups. These are: *Cereal and cereal substitutes; Pulses and pulse products; Milk and milk products; Edible oil, fruits, egg, fish and meat; Vegetables; Sugar, salt and spices; Beverages, pan, tobacco and intoxicants; Fuel and light; Clothing; Bedding and footwear* and *Miscellaneous non-food.* The demand system estimation requires price estimates for each of these consumption groups, separately for every unit in the analysis. We obtain these prices by calculating household-specific unit values directly from the NSS data. This is possible since the surveys include information on quantities and expenditure for the different consumption items. In all, we are able obtain such estimates for 155 consumption items.[12] Having obtained the household level unit values, we compute the median unit value within each unit. We next aggregate to the 11 consumption groups using the weighted country-product-dummy method (WCPD) due to Rao (1990).[13] We

[8] The price indices are defined as follows: $\ln \alpha(p^j) \equiv \alpha_0 + \sum_i \alpha_i \ln p_{ij} + \frac{1}{2} \sum_i \sum_h \gamma_{ih} \ln p_{ij} \ln p_{hj}$ and $\ln \beta(p^j) \equiv \sum_i \beta_i \ln p_{ij}$.

[9] $\lambda(p^j) \equiv \sum_i \lambda_i \ln p_{ij}$.

[10] This latest survey round was collected as an exception to this practice, due to severe droughts in 2009–10.

[11] According to the Indian Census, they account for 96 per cent of the population in 1991, 95 per cent in 2001 and 94 per cent in 2011. As Jharkhand and Chhattisgarh were carved out of Bihar and Madhya Pradesh in 2000, they do not appear in the household surveys before 2004–05. They do, however, appear as regions in Bihar and Madhya Pradesh such that it is possible to single them out. Therefore, we proceed by using the post-partition state boarders.

[12] We drop items that either do not appear in every survey round, or that are reported in incompatible units across survey rounds.

[13] This method is a modification of the unweighted version first suggested by Summers (1973).

provide more details on this aggregation in Appendix A.

Clearly, unit values are only proxies for prices. One advantage of using unit values in our setting is that they could be calculated from a large set of observations (in contrast to retail price estimates which are often based on fairly small samples). Another advantage is that the unit values are linked to actual transactions as opposed to price quotations. Still, one potential concern is quality differences in the reported consumption goods. Because of this concern, we provide a robustness check where we apply a quality adjustment. It is comforting that this alternative set of prices reveals the same results as the main estimation.

The last consumption group (*miscellaneous non-food*) consists of goods for which we are not able to compute unit values. This is because the NSS does not collect information on quantities for these items. If the consumption group was equally important for rich and poor households, we could reasonably have estimated our model without this group. However, the data clearly suggest that the budget share devoted to these non-food items increases with total expenditure.[14] Thus, the consumption group could potentially be an important source of cost-of-living differences between the rich and the poor. Therefore, we proceed in a similar manner as Deaton (2008) and impute prices using information from the official state- and sector-wise CPIs. These CPIs consist of several sub-indices, such that it is possible to construct an index for goods corresponding to our residual group. Yet, the CPIs cannot provide estimates of price *levels* across space, which we need to estimate our demand system. Because of this we proceed by setting the price level of miscellaneous non-food goods in the first time period equal to the price level of food items in the same state and sector. For later periods we impute prices such that we match the relative inflation rate vis-à-vis food items observed in the CPIs. Appendix A describes this procedure in more detail.

The NSS values consumption of goods at subsidized prices through the Public Distribution System (PDS) at the actual prices paid. The PDS is a public scheme centered on providing quotas of subsidized food grains (mainly rice and wheat) to eligible households. Because of the restrictions on quantity, the program is best seen as providing implicit income transfers (Khera, 2011; Dreze and Khera, 2013; Himanshu and Sen, 2013). In the analysis we therefore value consumption of PDS rice and wheat at the unit-wise market prices.[15] Because of this we do not use the PDS unit values for rice and wheat in the calculation of aggregate prices for cereals. We discuss the adjustment for the PDS and how it affects our findings in more detail in Section 5.4.

[14] As there is significant consumption growth during over study period, the importance of the non-food group is also likely to change over time. The average budget share of miscellaneous non-food increases from 17 per cent in 1993-94 to 25 per cent in 2011-12 in the rural sector, and from 17 per cent to 24 per cent in the urban sector.

[15] The PDS items are likely to be very similar to the corresponding market goods. Our valuation is justified if households can re-sell the PDS items at market prices without any cost, or if the amount supplied by the program is below their desirable level. This later condition seems reasonable, given the fact that most households with PDS consumption also purchase additional quantities of the same goods in the regular market (see Columns (9) and (10) in Table 12).

3.2 Estimation of demand system

We estimate the 11 goods QUAIDS demand system based on the budget share formulation shown in Equation (10). In the estimation, we use data on 100 expenditure level groups from every unit (mean per capita expenditure and budget shares for each group), and a Seemingly Unrelated Regressions system (SUR) estimated by Maximum Likelihood. By using groups instead of individual household data, we implicitly assume that preferences are homothetic within each of the expenditure groups. We consider that the within group variance in total expenditure is small enough such that this aggregation is unproblematic. Moreover, the assumption of normally distributed error terms is more likely to hold with grouped data (Aasness and Rødseth, 1983).

We impose homogeneity and negativity of the substitution matrix in the estimation. The homogeneity restriction is imposed simply by excluding the 11th budget share equation and by normalizing all prices relative to this last consumption group. The negativity restriction on the Slutsky matrix is more challenging. We follow an approach first suggested by Lau (1978) and later applied by Moschini (1998), which is based on imposing negativity at a *single* data point.[16] Like Neary (2004), we impose negativity at the sample means. By an appropriate scaling of the data, the substitution terms in the Slutsky matrix at this point reduce to a simple function of parameters only.[17] Finally, we do not directly estimate on the Slutsky matrix, but rather on the Cholesky decomposition of its mean values.

Yet, even after imposing these restrictions, there are still 85 parameters to be estimated, most of them appearing in every budget share equation. We follow Blundell *et al.* (1999) in estimating the parameters in an iterative manner. This is done by putting restrictions on the price responsiveness in the demand system, setting the last n-k-1 rows of the Cholesky decomposition equal to zero. This gives a "semi-flexible" system of rank k, with a smaller number of parameters to be estimated. We gradually increase the allowed price responsiveness by increasing the rank, using the estimated coefficients from the preceding values of k as starting values. We keep increasing the rank until the likelihood function no longer improves, which happens at $k = 8$.

To obtain elasticities we first differentiate 10 and obtain:

$$\mu_i = \frac{\partial \omega_i}{\partial \ln y} = \beta_i + \frac{2\lambda_i}{\beta(p)} \ln y, \tag{12}$$

and then calculate the budget elasticity as:

$$e_i = \frac{\mu_i}{\omega_i} + 1. \tag{13}$$

[16] Thus, we cannot be certain that the restriction holds everywhere. It is more likely to be violated in points far away from where negativity was imposed.

[17] See Appendix C in Neary (2004) for a discussion on this.

Table 1 presents estimates for two of the key parameters in these expressions. Standard errors, derived through bootstrapping, are shown in parentheses.[18]

Since the budget share equations are non-linear, the elasticities will vary with total expenditure. From the table it could still be seen that *cereal and cereal substitutes* and *miscellaneous non-food* are the two consumption groups for which the budget shares vary the most with total expenditure. The budget share for cereals falls in total expenditure—at least for low levels of expenditure—whereas the budget share for miscellaneous non-food increases for all expenditure levels.

[Table 1 about here]

4 Findings

The estimation procedure described above provides all parameters needed to compute the expenditure function given in Equation (11). This, combined with consumption group prices, is sufficient to calculate cost-of-living and real consumption inequality.[19]

Table 2 displays population weighted all-India cost-of-living measures by the rural and the urban sector, relative to the first time period. The differences across the consumption index and the two expenditure indices are fairly small for this aggregated statistic. However, the aggregated numbers mask important differences across households *within* units. One illustration of this can be found in Figure 1. The numbers underlying the figure are derived by first comparing the cost-of-living for households in the bottom and upper two expenditure percentiles relative to the average in each unit. Then, we display the average over all such comparisons, separately for the rural and the urban sector. Since the figure measures relative increases in cost-of-living, a number above (below) unity indicates that households in the particular expenditure group experienced a higher (lower) increase in the cost-of-living as compared to the average household. The figure suggests that the period from 1993–94 to 2004–05 was pro-poor, in the sense that the cost-of-living increased relatively more for the rich than for the poor. Whereas the cost-of-living rose by almost 100 per cent on average for the richest one per cent in each unit, it rose by roughly 80 per cent on average for the one per cent poorest, in this period. The overall relative price changes in the

[18]We conduct the bootstrapping as follows. We start with the sample of 100 expenditure groups for each unit. Then, we draw observations from this sample, with replacement, such that we match the original number of observations. We do this 1000 times, and estimate the demand system for each of the new samples. Finally, we construct standard errors using these 1000 sets of parameter estimates.

[19]We produce standard errors for both of these. For this purpose, we use the original data sample and the set of estimated demand parameters from the bootstrapping procedure to compute 1000 different estimates of the particular statistic. Hence, the standard errors derived over these different estimates capture the uncertainty related to the estimated demand model. As the number of observations in the data set is large, the coefficients of the demand system are precisely estimated and the standard errors are generally very small. This means that almost all differences we see, across price indices and across the subsequent inequality measures, are statistically significant (p-values< 0.00).

period thereafter are pro-rich, however, and hence the effect is somewhat dampened if considering the whole period up until 2011-12.

[Table 2 about here]

[Figure 1 about here]

We now proceed to investigate the full expenditure distribution, by computing inequality estimates. In this section, we focus on one particular measure, namely the Theil index. In Table 3, we present two other standard inequality measures, the Gini index and mean relative deviation, and show that our main findings are robust to the use of these alternative measures. We also present standard errors for the different inequality numbers that rely on the estimated demand system, as well as the inequality measures broken down to state and urban and rural areas (Table 4). Note also that we use household-specific expenditure levels when calculating inequality (and not expenditure group aggregates).[20] Figure 2 displays trends in consumption inequality, measured by the Theil index. The first column in the figure presents inequality numbers for the rural and urban sample combined, whereas the second and third columns show inequality estimates for the two sectors separately. Note that the trends for all inequality estimates are significantly different from each other (the p-values for the difference in difference estimates are very small, p-value$<$ 0.00).

The consumption and the expenditure homothetic cost-of-living numbers reveal close to similar inequality estimates for all three samples.[21] Thus, allowing for substitution in consumption does not seem to be of any quantitative importance in this application (although the differences are statistically significant, p-value$<$ 0.00). The expenditure non-homothetic estimates deviate more substantially. In particular, these estimates suggest a more moderate increase in inequality over the period 1993–94 to 2004–05, indicating a pro-poor development in cost-of-living (the difference-in-difference estimates are statistically significant, p-value$<$ 0.00). For the next five-year period, the opposite is true, and the homothetic indices underestimate the increase in inequality (again, p-value$<$ 0.00). This is especially noticeable in the rural sector where these estimates suggest a decrease in inequality, whereas the estimates that allow for non-homotheticity reveal a modest increase.

One advantage of the Theil index is that it is easily decomposable. The bottom panel of Figure

[20] We use the 17 major states when calculating inequality. Before computing the inequality estimates, we remove the 0.1 per cent poorest and the 0.1 per cent richest households within each unit. This exclusion is done because we are afraid that some of the extreme outliers are due to measurement errors. Our main findings are invariant to the inclusion/exclusion of these households.

[21] Note that the NSS survey from 1999-00 is not fully compatible with the other survey rounds, due to some inconsistencies in the recall periods used. See Deaton and Kozel (2005) for a detailed discussion on this. The level of inequality in 1999-00 might therefore not be comparable with the levels in the other years. Still, we have no reasons to expect that the inconsistency in recall period affects the differences between our three real expenditure measures.

2 displays between-group inequality estimates. The general pattern shown in the graphs suggests that inequalities in average expenditure levels between the rural and the urban sector, between states, and between states and sectors, have all increased over the period studied. Moreover, the figures clearly suggest that the differences between the non-homothetic and the other real expenditure measures do not stem from any of these between-group dimensions, but rather from the fact that the former index adjusts for cost-of-living differences within units.

[Figure 2 about here]

4.1 Discussion

It turns out that we can explain a large fraction of the difference between the homothetic and the non-homothetic inequality estimates by changes in the relative prices of *cereals* versus *miscellaneous non-foods*. As can be seen from Table 1, these two consumption groups are the ones for which the budget shares change the most with total consumption: the budget share of cereals decreases as households become richer, whereas the budget share of miscellaneous non-food increases. Figure 3 plots the percentage changes in the cereals/miscellaneous non-food price ratio along with the percentage changes in the ratio of the non-homothetic Theil index and the homothetic Theil index. As can be seen, periods in which the relative price ratio decreases (increases) are generally accompanied by a decrease (increase) also in the ratio of inequality estimates—meaning that the homothetic estimates overvalue (undervalue) inequality. We find the same pattern for the inequality estimates at the state level. These correlations are shown in terms of regression coefficients in Table 5.[22]

[Figure 3 about here]

[Table 5 about here]

These patterns illustrate how conventional measures overstate real inequality in situations when the prices of luxury goods increase relative to the prices of necessities, and vice versa when the opposite is the case. This is the case since these measures ignore the fact that relative price changes have a different effect on people at different income levels. However, relative price changes are likely to affect consumption inequality, not only through this cost-of-living but also through other channels, in particular through people's income. This latter channel is present in all measures of consumption inequality—without any particular adjustments—since income is likely to be reflected in expenditure data. The fact that the conventional measures incorporate the income effect but not the effect of the change in cost-of-living, may lead to a systematic bias.

[22] In order to compare changes over equally long time spells, we exclude the shorter time span between the two latest survey rounds.

The non-homothetic inequality numbers, especially for the rural sector, exhibit less variation over time as compared to the similar homothetic inequality estimates. This is consistent with a hypothesis of segregation in consumption/production. By this we mean that individuals' consumption patterns are skewed towards goods intensively produced by people of their own income group. One stark example of such segregation would be the dual economy, dating back to Lewis (1954). In a simple version of the Lewis model, there is one poor sector producing basic subsistence consumption goods, and people employed in this sector consume these goods and little else. There is also another rich and sophisticated sector producing a variety of goods and services, largely for those employed in the latter sector. Such a stark segregation is not very likely in reality, but there might very well be *some degree* of segregation.[23] If this is the case, then the conventional measures that rely on homothetic cost-of-living adjustments will tend to exaggerate changes in inequality that are due to relative price changes.

Below we discuss a simple theoretical framework to illustrate how segregation affects cost-of-living and real consumption inequality. Then, we present some indicative evidence for the segregation hypothesis by investigating rural crop producers.

4.1.1 Segregation, cost-of-living, and inequality: A simple theoretical framework

Suppose that we have two consumption goods: food (x_f) and non-food (x_{nf}), and two agents: rich (r) and deprived (d). To simplify the exposition, suppose further that preferences could be characterized by the following quasi-linear utility function:

$$U(x_f, x_{nf}) = ln(x_f) + x_{nf}. \tag{14}$$

Agents maximise utility subject to the budget constraint $p_f x_f + p_{nf} x_{nf} = w^i$, where w^i denotes total income for agent i, while p_f and p_{nf} denote the price of food and non-food, respectively. This simple maximization problem yields the following income-dependent budget share equations:

$$S_f = \frac{p_{nf}}{w^i}, \quad S_{nf} = 1 - \frac{p_{nf}}{w^i}. \tag{15}$$

From this, we see that the budget share spent on the non-food good, S_{nf}, increases in income. Moreover, the indirect utility function following from the maximization problem takes the form:

$$V(p, w^i) = ln\left(\frac{p_{nf}}{p_f}\right) + \left(\frac{w^i}{p_{nf}} - 1\right). \tag{16}$$

Suppose now that the price of the non-food good, for some exogenous reason, increases from p_{nf}^0 to p_{nf}^1 (whereas the price of the food good, p_f, remains constant). We denote the vector of prices

[23] See Ray (2013) for a brief discussion on this type of segregation.

as p^0 and p^1. The individual-specific expenditure non-homothetic index, comparing period 1 with period 0, now takes the form:

$$P_1^i = \frac{e(p^1, v(p^1, w^i))}{e(p^0, v(p^1, w^i))} = \frac{w^i}{\left(ln(\frac{p_{nf}^1}{p_f}) - ln(\frac{p_{nf}^0}{p_f}) + \frac{w^i}{p_{nf}^1}\right) p_{nf}^0} = \frac{1}{\left(ln(\frac{p_{nf}^1}{p_f}) - ln(\frac{p_{nf}^0}{p_f})\right) \frac{p_{nf}^0}{w^i} + \frac{p_{nf}^0}{p_{nf}^1}} \tag{17}$$

Since $\partial P_1^i / \partial w_i > 0$ and $w^r > w^d$, we see that the cost-of-living increases more for the rich agent than for the deprived agent. This is intuitive, since the rich agent allocates a larger share of its budget towards the non-food good and is thus relatively more hurt by the price increase.

If we denote the average income level by $\bar{w}$, we could write the expenditure homethetic index comparing period 1 to period 0 as follows:

$$\bar{P}_1 = \frac{e(p^1, v(p^1, \bar{w}))}{e(p^0, v(p^1, \bar{w}))} = \frac{1}{\left(ln(\frac{p_{nf}^1}{p_f}) - ln(\frac{p_{nf}^0}{p_f})\right) \frac{p_{nf}^0}{\bar{w}} + \frac{p_{nf}^0}{p_{nf}^1}}. \tag{18}$$

Since $w^r > \bar{w} > w^d$, it must also follow that $P_1^r > \bar{P}_1 > P_1^d$. Thus, the increase in cost-of-living, according to this measures, lies somewhere in-between the individual increases for the rich and the deprived agent, identified in (17). Note that $\bar{P}_1$ would generally not correspond to the average of P_1^r and P_1^d. Furthermore, real consumption inequality as measured by the non-homothetic index would be lower than its homothetic counterpart. If we measure inequality simply as the fraction of total consumption acquired by the rich agent, this could be seen by the following expression:

$$\frac{w^r}{(w^r + w^d)} = \frac{w^r / \bar{P}_1}{(w^r / \bar{P}_1 + w^d / \bar{P}_1)} > \frac{w^r / P_1^r}{(w^r / P_1^r + w^d / P_1^p)}. \tag{19}$$

Let us now introduce the production side. More particularly, let us assume that the deprived agent gets its labour income from the production of food, while the rich agent gets its income from the production of non-food items. Finally, let us assume that both agents are paid according to their marginal productivity. That is:

$$w^d = p_f F_L, \quad w^r = p_{nf} G_L, \tag{20}$$

where F_L and G_L denote the marginal productivity of labour in food and non-food production, respectively. From this assumption, it follows that the increase in the price of the non-food good increases the labour income of the rich agent. Nominal inequality will therefore also rise. Since the homothetic index uses the same cost-of-living deflator for both agents, it would give the exact same inequality levels as the nominal values.

We have already seen that the price increase affects the cost-of-living of the rich agent more severely than the cost-of-living of the deprived agent. Thus, the demand side effect dampens the increase in inequality:

$$\frac{w_0^r}{(w_0^r + w_0^d)} < \frac{w_1^r / P_1^r}{(w_1^r / P_1^r + w_1^d / P_1^d)} < \frac{w_1^r}{(w_1^r + w_1^d)}. \tag{21}$$

This stylized illustration shows how segregation in consumption/production dampens changes in consumption inequality that are due to relative price changes. The next subsection provides some empirical evidence on the presence of such segregation based on comparing crop producers with other households.

4.1.2 Segregation, cost-of-living, and inequality: An empirical investigation

Table 6 presents some background statistics. As can be seen, more than half of the rural households are engaged in crop production and those engaged are, on average, poorer than households in other activities. The bottom part of the table shows similar numbers for rural food crop labours (these numbers exclude self-employed crop producers). As revealed from the consumption ratios, these households are considerably poorer: they are, on average, roughly 25 per cent poorer than rural households outside crop production, and about 30 to 40 per cent poorer than all households outside crop production.

When crop prices rise more than other prices, we would expect the nominal incomes of these rural farmers, and hence also their total expenditure, to increase relative to the income and expenditure of other households. To investigate this, we look at the correlation between the percentage changes in the expenditure ratios (from Table 6) and the percentage changes in the price of cereals over the price of miscellaneous non-foods. Table 7 presents regression estimates at the state level. The estimates shown in the first two columns are based on expenditure ratios for all crop producers, whereas those in the third and the fourth column are based on crop labourers only. As can be seen from the positive coefficients, periods with relatively large (small) increases in the prices for cereals are accompanied by an increase (decrease) in this expenditure ratio, meaning that the gap between crop labourers and other labourers decreases (increases). This suggests that their labour income changes when there is a change in relative prices. As the crop producers are relatively poor, we would also expect that the income effect helps decrease overall inequality. Columns (1) and (2) in Table 8 present the coefficients from two regressions of the percentage changes in inequality versus the percentage changes in the relative price of cereals versus miscellaneous non-foods, by states. Not surprisingly, increases in these relative prices are associated with lower consumption inequality, as measured by the homothetic measures. This holds within the rural sector, and more strongly for the rural and the urban sector combined.

Our non-homothetic index captures this nominal income effect but, in addition, it captures the differential effect of price changes on cost-of-living. Given that the crop producers are relatively poor, and given that the budget share of cereals tends to fall with total expenditure, we would expect the group as a whole to devote a higher-than-average share of its budget towards cereals. In Table 9, we show estimates from two regressions of cereal budget shares versus two binary variables indicating whether a household is engaged in crop production as labour or self-employed, respectively. The regression coefficients shown in the first column suggest that the crop labourers allocated almost a five percentage point higher budget share towards cereals as compared to other households outside crop production. In the regression shown in the second column, we include the logarithm of per capita expenditure and its square as explanatory variables. When doing this, the size of the coefficient decreases by more than two-thirds, indicating that the main reason for the relatively high cereal budget share is that these households are poor.

Since the farmers devote a relatively large share of their budget towards cereals, higher cereal prices will have a relatively large impact on their cost-of-living, which according to the theoretical exposition should weaken the relationship between changes in relative cereal price and inequality. As can be seen from Columns (3) and (4), in Table 8, the association is still negative, but the size of the coefficients is of a small magnitude and they are not statistically significantly different from zero.

What this empirical exercise illustrates is that the conventional inequality measures tend to overestimate changes in inequality that are due to changes in relative prices when there is segregation in production/consumption.

[Table 6 to 9 about here]

5 Robustness

In this section, we present four types of robustness checks. All these alternative specifications provide similar trends in real consumption inequality as in our main analysis. Moreover, for all specifications, we find that the allowance for non-homotheticity is quantitatively much more important than the allowance for substitution in consumption.

5.1 Alternative references prices

As a first robustness check, we compute the cost-of-living indices using two alternative sets of reference prices. First, we adopt the procedure suggested by Barnett *et al.* (2009), and later implemented by Feenstra *et al.* (2012). This procedure is based on using *every* unit's price vector as a reference, and then taking a geometric mean of all such comparisons. For brevity, we refer to these references as "Diewert prices". Using the Diewert prices as a base price vector, the real consumption level of unit j derived through the consumption index could be expressed as:

$$I_j^{cons} = \prod_s^m (p^s q^j)^{\frac{1}{m}}. \tag{22}$$

The expenditure homothetic index becomes:

$$I_j^{exp\text{-}h} = \prod_s^m e(p^s, v(p^j, z_j))^{\frac{1}{m}}, \tag{23}$$

whereas the expenditure non-homothetic index can be written as:

$$I_j^{exp\text{-}nh} = \prod_s^m \left(L^{-1} \sum_l e(p^s, v(p^j, z_{jl})) \right)^{\frac{1}{m}}. \tag{24}$$

As a second set of alternative reference prices, we simply use all unit prices as references, instead of taking the geometric mean. As most methods of calculating reference prices would produce some average of the price vectors of the individual units, this procedure should be seen as extremely flexible. However, for most applications, it is not very convenient, as it gives the same number of real consumption estimates for each unit as the total number of units.

Figure 4 plots the trends in inequality using these different reference price vectors. The left column shows the expenditure non-homothetic Theil index, whereas the middle and the right columns plot the difference between these numbers and the inequality estimates derived through the consumption index and the expenditure homothetic index, respectively.[24] The solid lines, labeled "Geary ref.", are based on the Geary reference prices (as the inequality estimates presented in the main analysis), while the dotted lines, labeled "Diewert ref.", are based on the cost-of-living measures using the Diewert reference prices. Finally, the light grey lines use the price vectors of all units as references. As could be seen from all three panels, the choice between the Geary and the Diewert reference prices does not affect the subsequent inequality estimates (they are indistinguishable in the graphs). We get somewhat different inequality numbers within the large set of reference prices based on the price vectors of every unit, but as can be seen from the grey bands, the trends in both inequality levels as well as in the difference between the inequality measures are not affected to any considerable extent.

[Figure 4 about here]

[24]When doing this, we normalize the difference to zero in the first time period.

5.2 Quality-adjusted unit values

In the main analysis, we use median unit values as proxies for prices. Even though we are able to compute these unit values at a fine level of goods disaggregation, many of the consumption items might still not be perfectly homogeneous. This could be problematic, as households' reported unit values will be affected by the quality of the underlying good. If households from different regions systematically purchase goods with different quality levels, then median unit values will provide biased estimates of the true price differences. Deaton *et al.* (2004) suggest a regression-based method to correct for this possible bias. They start out by assuming that variation in the reported unit values stems from a mixture of quality and true price differences:

$$ln\ uv_{il} = ln\ p_{ij} + ln\ \varphi_{il}, \tag{25}$$

where uv_{il} is the unit value of item i reported by household l, p_{ij} is the true item price in unit j (at some base quality level common for every unit), while φ_{il} is the quality of the item consumed by household l. A convenient assumption is that the chosen quality could be represented as a log-linear function of real consumption:

$$ln\ uv_{il} = ln\ p_{ij} + b_i ln\ y_l + \gamma X, \tag{26}$$

where y_l is the real consumption level of household l, and X is a vector of other possible household covariates. The b_i-coefficient could be interpreted as the elasticity of quality with respect to total expenditure. From this it could be seen that the quality-bias in the unit values is a function of the real consumption level and the quality elasticity. The procedure proposed in Deaton *et al.* (2004) only partially removes this bias, since it replaces real per capita expenditure with nominal per capita expenditure. Provided that cost-of-living differs across regions and over time, the quality-adjusted prices will therefore also include a bias, which depends on the expenditure elasticity and the overall price level in each unit. More particularly, the estimated item prices in a unit would be more biased the further apart the cost-of-living in the unit is from the average. Provided that the expenditure elasticity is positive, meaning that the quality consumed increases with total expenditure, we could also infer that the procedure underestimates spatial cost-of-living differences across units, as it undervalues the item prices in high-cost areas and overvalues the item prices in low-cost areas. By the same logic, we could infer that the procedure underestimates increases in cost-of-living over time—provided that the overall cost-of-living rises—since it overestimates item prices in early time periods, and underestimates item prices in later time periods.

The bias could be avoided by replacing nominal expenditure in Equation (26) by real expenditure. The main challenge is that we need the unbiased item prices to derive an estimate of the overall cost-of-living in each unit. Therefore, we propose an iterative method. In the first step, we estimate the following regression, separately for every item i, using nominal per capita expenditure values as in Deaton *et al.* (2004):

$$ln\ uv_{il} = \sum_j d_j D_j + b ln z_{lj} + \gamma X, \quad (27)$$

where D_j is a set of unit dummies, z_{lj} is the nominal expenditure level of household l living in unit j and X is a vector of household covariates (the number of household member below 16 years old, the number of members above 16 and the age of the household head). We identify the price component from the unit dummies. The bias in the subsequent price measure of item i can be expressed as:

$$ln\ p_{ij} - ln\ \hat{p_{ij,1}} = b_i ln(\overline{e(\pi, v(p^j, z_{jl}))}) - \hat{b}_i ln \overline{z_{lj}}, \quad (28)$$

where $\overline{e(\pi, v(p^j, z_{jl}))}$ and $\overline{z_{lj}}$ display the mean real and nominal expenditure levels in unit j, respectively, relative to some base. The subscript on $\hat{p}_{ij,1}$ denotes that this is our first estimate of p_{ij}. Next, we use these proxies of the item prices to estimate aggregated consumption group prices, and then to compute our non-homothetic cost-of-living index as described in Section 2. Having obtained these overall cost-of-living measures, we re-run the regression from Equation (27), again separately for every item i, but now using the real expenditure measures instead of the nominal values:

$$ln\ uv_{il} = \sum_j d_j D_j + b ln(e(\hat{\pi}_1, v(\hat{p^j_1}, z_{jl}))) + \gamma X. \quad (29)$$

From this estimation, we are able to extract a new set of item price measures. The bias in this price estimate of item i can be expressed as:

$$ln\ p_{ij} - ln\ \hat{p_{ij,2}} = b_i ln(\overline{e(\pi, v(p^j, z_{jl}))}) - \hat{b}_i ln(\overline{e(\hat{\pi}_1, v(\hat{p^j_1}, z_{jl}))}). \quad (30)$$

The absolute size of the bias in $ln\ \hat{p}_{ij,2}$ is smaller than the bias in $ln\ \hat{p}_{ij,1}$, provided that:

$$\left| b_i ln(\overline{e(\pi, v(p^j, z_{jl}))}) - \hat{b}_i ln(\overline{e(\hat{\pi}_1, v(\hat{p^j_1}, z_{jl}))}) \right| < \left| b_i ln(\overline{e(\pi, v(p^j, z_{jl}))}) - \hat{b}_i ln \overline{z_{lj}} \right|. \quad (31)$$

Hence, if this requirement is fulfilled, we could repeat the procedure and the solution should eventually converge.

Table 10 presents unit value estimates for the eight most important items in terms of average budget shares. All numbers in the table are derived as population weighted averages of the unit specific numbers. The first row for each good shows the median unit values (that is, averages over the median unit values within each unit), whereas the second row presents quality adjusted numbers based on the methodology in Deaton *et al.* (2004). The following five rows show the unit value estimates from the five succeeding iterations in our proposed procedure. The numbers in parenthesis display the b-coefficients from the item-specific regressions. If each of these item groups had consisted of completely homogeneous goods, these coefficients should have been close to zero. For items such as sugar and oils, which are likely to be rather homogeneous, we see that the coefficients are indeed close to zero. Thus, the biases in the median unit values and in the adjusted prices from the methodology in Deaton *et al.* (2004) are likely to be small. However, goods within consumption headings such as "garments" are clearly more heterogeneous, and the two aforementioned procedures are therefore likely to produce more seriously biased price estimates.

[Table 10 about here]

Figure 5 presents the price trends for the different groups of consumption items.[25] A first thing to note from the figure is that the adjustment of Deaton *et al.* (2004) gives lower price increases as compared to the median unit values. This is as expected, given positive values of the different b-coefficients and increases in overall cost-of-living over time. A second thing to notice is that the price estimates derived from our iteration procedure are generally somewhere in-between the two other price estimates, although much closer to the median unit values. Therefore, it is not very surprising that the quality adjustment does not change our final inequality estimates to any considerable extent. The graphs to the left in Figure 6 show inequality trends derived from our non-homothetic cost-of-living index, separately for the median unit values and the quality adjusted unit values, whereas the middle and the right graphs display the difference between these numbers and the inequality numbers from the consumption and the expenditure homothetic indices, respectively. As can be seen, the differences between the two sets of estimates are close to negligible.

[Figure 5 about here]

[Figure 6 about here]

5.3 Equivalence scaling and demographics

As a third robustness check, we repeat the whole analysis using equivalence scaling. The key difference between these estimates and those in the main analysis is the composition of households in the expenditure groups used for estimation of the demand system and for the calculation of the cost-of-living indices. Various equivalence scales have been proposed in the literature. We use the standard OECD scale of 1982. 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 an individual aged below 16.

The resulting inequality estimates are presented in Figure 7. The use of equivalence scales reduces the levels of inequality somewhat, as can be seen from the graph in the left column. Still, the trends in inequality, as well as the differences between the various estimates, are almost identical to the main analysis.

[25] We do not present the 11th group here, since it is derived by combining the unit value estimates with information from the official CPIs.

[Figure 7 about here]

Relative prices may affect people differently not only because preferences are non-homothetic, but also because people live in households with different compositions. To more directly test whether our results are driven by differences in family composition, we conduct the whole analysis for a range of subsamples. In each of these subsamples, we only include households with an identical composition of adults and children.

Table 11 shows the number of households in each of our subsamples. Since we need a reasonable number of households within each state, we choose subsamples with at least 3000 observations in each survey round. Still, there are too few observations within each of these to construct percentiles for every unit. Therefore, we base the estimation of the QUAIDS system, and the subsequent cost-of-living measures, on 20 expenditure groups instead of 100 as in the main analysis.

Figure 8 and 9 display cost-of-living for the bottom and upper two expenditure groups, relative to the average, for the rural and the urban sector, respectively. As can be seen from these figures, the trends are very similar across all these subsamples, which suggests that our inequality and cost-of-living trends are not driven by differences in family composition.[26]

[Table 11 about here]

[Figure 8 and 9 about here]

5.4 The Public Distribution System (PDS)

In the main analysis, we value the consumption of subsidized goods through the PDS at local market prices. As a fourth robustness check, we here estimate cost-of-living and inequality while evaluating these goods at the actual prices paid. This robustness check is interesting in its own right, as it tells us something about the distributional impact of the public scheme.

Table 12 presents some background statistics of the PDS. The first two columns show the share of households consuming any PDS rice and PDS wheat, respectively, while the next two columns display the average per capita quantities consumed among these households. As can be seen, the average quantities are fairly stable over time, while the coverage of households—especially in the rural sector—has increased substantially. Columns (5) to (8) display the average across the median unit values for the subsidized PDS items and the corresponding market items. The PDS prices have been close to constant over time, whereas market prices have grown roughly threefold—meaning that the value of having access to the scheme has risen substantially over

[26] The inequality numbers for each of these subsamples are less interesting, since they are based on completely different populations than those in our main analysis.

time. This, together with the increase in coverage, means that the choice of how to treat PDS consumption will be more important for the later survey rounds. The two final columns present the fraction of households with PDS consumption of either rice or wheat that also consume the same goods from the regular market. As can be seen from these columns, the majority of the PDS households purchase additional amounts from the regular market.

Figure 10 shows how the inequality estimates change when we evaluate the PDS items at actual prices paid. As the program is (at least intentionally) targeted towards the poor, it is not surprising that the inequality numbers rise somewhat as compared to those presented in the main analysis.[27] However, the trends and the differences between the three sets of estimates are similar.

[Table 12 about here]

[Figure 10 about here]

6 Concluding remarks

This paper studies price changes and real consumption inequality in India through adjusting for expenditure specific cost-of-living. We find that in periods when the prices of necessities decreased relative to other goods (1993-94 to 2004-05 and 2009-10 to 2011-12), traditional indices overestimate the increase in inequality whereas the opposite is true for the period when the prices of necessities increased relative to other goods (2004-05 to 2009-10). We also show that taking non-homotheticity into account is quantitatively much more important than the adjustment for substitution in consumption, despite the larger attention given to the substitution bias in the price index literature.

All of our findings are robust to various robustness checks. Moreover, our inequality estimates exhibit fewer fluctuations over time as compared to estimates that do not allow for non-homotheticity. This difference is consistent with a hypothesis of segregation in consumption/production, meaning that people intensively consume goods produced by people of their own income group. If such segregation is present, we would also expect conventional estimates to overstate changes in inequality due to relative price changes. We provide some empirical support for this hypothesis, by comparing households in crop production with other types of households.

[27] For evidence on how the PDS affects measures of poverty, see Dreze and Khera (2013) and Himanshu and Sen (2013).

References

Aasness, J. and Rødseth, A. (1983). Engel Curves and Systems of Demand Functions. *European Economic Review*, **20**(1), 95–121.

Akmal, M. and Dowrick, S. (2005). Contradictory Trends in Global Income Inequality: A Tale of Two Biases. *Review of Income and Wealth*, **51**(2), 201–230.

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**.

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

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

Barnett, W. A., Diewert, W. E., and Zellner, A. (2009). Introduction to Measurement With Theory. *Macroeconomic Dynamics*, **13**(S2), 151–168.

Blundell, R., Robin, J. M., *et al.* (1999). Estimation in Large and Dissaggregated Demand Systems: An Estimator for Conditionally Linear Systems. *Journal of Applied Econometrics*, **14**(3), 209–232.

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.

Cravino, J. and Levchenko, A. A. (2015). The Distributional Consequences of Large Devaluations.

Crossley, T. F. and Pendakur, K. (2010). The Common-Scaling Social Cost-of-Living Index. *Journal of Business & Economic Statistics*, **28**(4), 523–538.

Datt, G. and Ravallion, M. (2009). *Has India's Economic Growth Become More Pro-poor in the Wake of Economic Reforms?* World Bank Publications.

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

Deaton, A. and Dupriez, O. (2011). Spatial Price Differences Within Large Countries.

Deaton, A. and Kozel, V. (2005). Data and Dogma: The Great Indian Poverty Debate. *The World Bank Research Observer*, **20**(2), 177–199.

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

Deaton, A., Friedman, J., and Alatas, V. (2004). Purchasing Power Parity Exchange Rates from Household Survey Data: India and Indonesia. *Princeton Research Program in Development Studies Working Paper*.

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

Dreze, J. and Khera, R. (2013). Rural Poverty and The Public Distribution System. *Economic & Political Weekly*, **48**(45-46), 55–60.

Faber, B. (2012). Trade Liberalization, the Price of Quality, and Inequality: Evidence from Mexican Store Prices.

Feenstra, R. C., Ma, H., Peter Neary, J., and Prasada Rao, D. (2012). Who Shrunk China? Puzzles in the Measurement of Real GDP. *The Economic Journal*, **123**(573), 1100–1129.

Feenstra, Robert C., R. I. and Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review*, **105**(10), 3150–3180.

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

Handbury, J. (2013). Are Poor Cities Cheap for Everyone? Non-homotheticity and the Cost of Living Across US Cities. *The Wharton School Research Paper*, (71).

Himanshu and Sen, A. (2013). In-Kind Food Transfers: Impact on Poverty. *Economic & Political Weekly*, **48**(45-46), 46–54.

Khera, R. (2011). India's Public Distribution System: Utilisation and Impact. *Journal of Development Studies*, **47**(7), 1038–1060.

Lau, L. J. (1978). Testing and imposing monoticity, convexity and quasi-convexity constraints. *Production economics: A dual approach to theory and applications*, **1**, 409–453.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, **22**(2), 139–191.

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

Moretti, E. (2013). Real Wage Inequality. *American Economic Journal: Applied Economics*, **5**(1), 65–103.

Moschini, G. (1998). The Semiflexible Almost Ideal Demand System. *European Economic Review*, **42**(2), 349–364.

Muellbauer, J. (1974). Prices and Inequality: The United Kingdom Experience. *The Economic Journal*, pages 32–55.

Muellbauer, J. (1976). Community Preferences and the Representative Consumer. *Econometrica*, **44**(5), 979–999.

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.

Nicholas, A., Ray, R., and Valenzuela, M. R. (2010). Evaluating the Distributional Implications of Price Movements: Methodology, Application and Australian Evidence. *Economic Record*, **86**(274), 352–366.

Pendakur, K. (2002). Taking Prices Seriously in the Measurement of Inequality. *Journal of Public Economics*, **86**(1), 47–69.

Pollak, R. A. (1980). Group Cost-of-Living Indexes. *The American Economic Review*, **70**(2), 273–278.

Pollak, R. A. (1981). The Social Cost of Living Index. *Journal of Public Economics*, **15**(3), 311–336.

Rao, D. (1990). A System of Log-Change Index Numbers for Multilateral Comparisons. In *Comparisons of Prices and Real Products in Latin America*, pages 127–139. New York: Elsevier Science Publishers.

Ray, D. (2013). Notes for a Course in Development Economics.

Summers, R. (1973). International Price Comparisons Based Upon Incomplete Data. *Review of Income and Wealth*, **19**(1), 1–16.

The World Bank (2011). *Perspectives on Poverty in India: Stylized Facts from Survey Data*. World Bank Publications.

Weisskopf, T. E. (2011). Why Worry About Inequality in the Booming Indian Economy. *Economic & Political Weekly*, **47**, 41–51.

Appendix A Aggregation and imputation of prices for miscellaneous non-food

This section explains in detail how we construct the consumption group price measures.

Having obtained the median unit values for each unit (states-sectors), we aggregate the item level estimates to consumption groups. This aggregation is done using the weighted country-product-dummy method (WCPD) due to Rao (1990). The procedure is based on a set of regressions, where the logarithm of the item prices is regressed on a set of dummy variables using weighted least squares. We thus run the following regression, separately for every consumption group:

$$ln\ \hat{p}_{ij} = \sum_j \alpha_j D_j + \sum_i b_i D_i, \tag{32}$$

where D_j is a set of unit dummy variables, and D_i is a dummy variable for every item i in each consumption group. We use the item-wise average budget shares in each unit as weights. Finally, the aggregate price estimates for each consumption group are found directly from the dummy coefficients as:

$$ln\ \hat{p}_j = \alpha_j. \tag{33}$$

The last consumption group (*miscellaneous non-food*) consists of goods for which we are not able to compute unit values, since the NSS does not collect information on quantities for these items. There is no straightforward way of imputing prices for this residual group. Yet, it seems most natural to use the official state-specific consumer price index (see also Deaton, 2008). We proceed as follows: we first calculate a unit-specific food price index using the price estimates from all food items and the WCPD method. The all-India values of this food price index are displayed in Table 13, relative to the first time period and separate for the rural and the urban sector. The second column in the table shows the corresponding numbers from the *Consumer Price Index for Industrial Workers* (CPIIW) for the urban sector, and the *Consumer Price Index for Agricultural Labourers* (CPIAL) for the rural sector. These numbers are derived as the weighted average of the state-specific indices. In the third column, we show the CPI sub-index that corresponds to our *miscellaneous non-food* consumption group.[28] Finally, the fourth column presents the CPI food price increases, relative to the miscellaneous non-food price increases. We use this ratio to scale our last consumption group. This seems like a reasonable procedure; especially since our unit value food indices roughly follow the same trends as the CPI food indices. For the first period, we set the prices of the *miscellaneous non-food* group in each unit equal to their values for the unit value food index. For later periods, we impute values equal to the same food index multiplied by the relative inflation rates displayed in the fourth column of the table.

[28]For the rural sector, this CPI sub-index exactly matches our 11th consumption group, while the CPI for the urban sector has two sub-indices for the goods included in our residual group. The sub-index displayed in the table is therefore constructed as a weighted average of the *miscellaneous non-food* and the *housing* CPI sub-indices.

[Table 13 about here]

Appendix B Tables and figures

Table 1: Parameters QUAIDS

	β		λ	
Cereal and cereal substitutes	-0.1224	(0.0011)	0.0308	(0.0014)
Pulses and pulse products	-0.0111	(0.0002)	-0.0024	(0.0002)
Milk and milk products	0.0201	(0.0009)	-0.0293	(0.0009)
Edible oil, fruits, egg, fish and meat	-0.0086	(0.0005)	-0.0105	(0.0005)
Vegetables	-0.0213	(0.0003)	-0.0008	(0.0006)
Sugar, salt and spices	-0.0136	(0.0002)	-0.0033	(0.0003)
Beverages, pan, tobacco and intoxicants	0.0150	(0.0007)	-0.0026	(0.0009)
Fuel and light	-0.0213	(0.0004)	-0.0030	(0.0008)
Clothing	-0.0083	(0.0002)	-0.0020	(0.0004)
Bedding and footwear	0.0021	(0.0001)	-0.0020	(0.0001)
Miscellaneous non-food	0.1693	(0.0010)	0.0252	(0.0020)

Note: The table displays two of the key parameters from the estimation of the QUAIDS demand system. Standard errors derived through bootstrapping are shown in parentheses.

Table 2: All-India cost-of-living relative to 1993–94

	Consumption Index	**Expenditure Index**			
		Homothetic		Non-homothetic	
	(1)	(2)		(3)	
Rural					
1993–94	100.0	100.0	(0.00)	100.0	(0.00)
1999–00	158.8	158.4	(0.03)	158.0	(0.02)
2004–05	184.6	184.7	(0.07)	183.5	(0.06)
2009–10	280.0	281.1	(0.07)	281.5	(0.06)
2011–12	329.0	329.9	(0.08)	328.5	(0.08)
Urban					
1993–94	100.0	100.0	(0.00)	100.0	(0.00)
1999–00	156.6	156.5	(0.02)	156.6	(0.01)
2004–05	188.9	186.6	(0.07)	185.6	(0.06)
2009–10	286.2	285.0	(0.06)	285.4	(0.04)
2011–12	341.4	340.0	(0.09)	338.1	(0.08)

Note: All numbers are population weighted, using the multipliers provided by the NSS. The non-homothetic indices are normalized such that they give the same cost-of-living for all expenditure groups *within* each unit in the first period. Standard errors derived through bootstrapping are shown in parentheses.

TABLE 3: Inequality estimates

	Theil			Gini			Mean relative deviation		
	Cons	Exp-h	Exp-nh	Cons	Exp-h	Exp-nh	Cons	Exp-h	Exp-nh
Combined									
1993-94	0.14572	0.14598	0.14592	0.28356	0.28381	0.28382	0.13032	0.13058	0.13057
		(0.00001)	(0.00001)		(0.00001)	(0.00001)		(0.00001)	(0.00001)
1999-00	0.15844	0.15892	0.15574	0.29524	0.29579	0.29283	0.14041	0.14093	0.13806
		(0.00003)	(0.00003)		(0.00002)	(0.00003)		(0.00002)	(0.00003)
2004-05	0.20176	0.20461	0.19283	0.32473	0.32700	0.31871	0.17069	0.17301	0.16415
		(0.00002)	(0.00009)		(0.00001)	(0.00006)		(0.00001)	(0.00006)
2009-10	0.20476	0.20628	0.20491	0.32588	0.32711	0.32662	0.17169	0.17299	0.17231
		(0.00002)	(0.00011)		(0.00001)	(0.00007)		(0.00002)	(0.00007)
2011-12	0.21499	0.21522	0.20785	0.33516	0.33531	0.32988	0.18152	0.18169	0.17558
		(0.00002)	(0.00007)		(0.00002)	(0.00005)		(0.00002)	(0.00006)
Rural									
1993-94	0.11730	0.11724	0.11745	0.25655	0.25657	0.25680	0.13032	0.13058	0.13057
		(0.00001)	(0.00001)		(0.00001)	(0.00001)		(0.00074)	(0.00074)
1999-00	0.11970	0.12026	0.11663	0.25997	0.26065	0.25679	0.14041	0.14093	0.13806
		(0.00001)	(0.00003)		(0.00002)	(0.00003)		(0.00099)	(0.00100)
2004-05	0.14814	0.14893	0.14100	0.27945	0.28025	0.27311	0.17069	0.17301	0.16415
		(0.00002)	(0.00005)		(0.00001)	(0.00005)		(0.00142)	(0.00134)
2009-10	0.14105	0.14068	0.14208	0.27500	0.27477	0.27549	0.17169	0.17299	0.17231
		(0.00004)	(0.00007)		(0.00003)	(0.00007)		(0.00161)	(0.00156)
2011-12	0.15344	0.15252	0.14805	0.28683	0.28599	0.28106	0.18152	0.18169	0.17558
		(0.00005)	(0.00008)		(0.00005)	(0.00008)		(0.00156)	(0.00151)
Urban									
1993-94	0.16850	0.16877	0.16857	0.30914	0.30941	0.30927	0.13032	0.13058	0.13057
		(0.00001)	(0.00001)		(0.00000)	(0.00000)		(0.00076)	(0.00076)
1999-00	0.18367	0.18382	0.18255	0.32404	0.32421	0.32356	0.14041	0.14093	0.13806
		(0.00001)	(0.00005)		(0.00001)	(0.00004)		(0.00088)	(0.00095)
2004-05	0.22988	0.23052	0.21488	0.35807	0.35863	0.34778	0.17069	0.17301	0.16415
		(0.00001)	(0.00014)		(0.00001)	(0.00008)		(0.00110)	(0.00098)
2009-10	0.24020	0.24022	0.23571	0.36257	0.36285	0.36025	0.17169	0.17299	0.17231
		(0.00002)	(0.00017)		(0.00001)	(0.00010)		(0.00127)	(0.00120)
2011-12	0.23985	0.24009	0.23017	0.36222	0.36263	0.35582	0.18152	0.18169	0.17558
		(0.00001)	(0.00009)		(0.00001)	(0.00006)		(0.00098)	(0.00092)

Note: "Cons" denotes the consumption index; "Exp-h" denotes the expenditure homothetic index, and "Exp-nh" denotes the expenditure non-homothetic index. The mean relative deviation inequality measure is derived as the difference in the logarithms of the arithmetic mean and the geometric mean. The numbers in parenthesis display standard deviations derived through bootstrapping. We derive these by running 1000 replications of the estimation of the QUAIDS model, and by computing the subsequent inequality measures for all these replications. This work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR).

Table 4: Theil index by state and sectors

	1993–94			1999–00			2004–05			2009–10			2011–12		
	Cons	Exp-h	Exp-nh	Cons	Exp-h	Exp-nh	Cons	Exp-h	Exp-nh	Cons	Exp-h	Exp-nh	Cons	Exp-h	Exp-nh
Rural															
Andhra Pradesh	0.110	0.110	0.110	0.096	0.096	0.093	0.122	0.122	0.119	0.127	0.127	0.129	0.101	0.101	0.096
Assam	0.049	0.049	0.049	0.068	0.068	0.070	0.061	0.061	0.056	0.079	0.079	0.080	0.079	0.079	0.080
Bihar	0.074	0.074	0.074	0.070	0.070	0.071	0.063	0.063	0.061	0.077	0.077	0.079	0.070	0.070	0.070
Chhattisgarh	0.070	0.070	0.070	0.095	0.095	0.088	0.125	0.125	0.115	0.080	0.080	0.080	0.094	0.094	0.091
Gujarat	0.079	0.079	0.079	0.092	0.092	0.090	0.115	0.115	0.108	0.111	0.111	0.110	0.114	0.114	0.110
Haryana	0.119	0.119	0.119	0.101	0.101	0.089	0.213	0.213	0.186	0.128	0.128	0.127	0.105	0.105	0.097
Jharkhand	0.079	0.079	0.079	0.096	0.096	0.098	0.076	0.076	0.070	0.071	0.071	0.073	0.083	0.083	0.086
Karnataka	0.097	0.097	0.097	0.100	0.100	0.101	0.114	0.114	0.108	0.092	0.092	0.099	0.133	0.133	0.133
Kerala	0.133	0.133	0.133	0.122	0.122	0.126	0.187	0.187	0.186	0.187	0.187	0.197	0.191	0.191	0.197
Madhya Pradesh	0.105	0.105	0.105	0.107	0.107	0.095	0.113	0.113	0.108	0.132	0.132	0.130	0.123	0.123	0.120
Maharashtra	0.124	0.124	0.124	0.115	0.115	0.118	0.146	0.146	0.141	0.101	0.101	0.101	0.120	0.120	0.109
Odisha	0.096	0.096	0.096	0.098	0.098	0.097	0.123	0.123	0.116	0.098	0.098	0.100	0.087	0.087	0.086
Punjab	0.099	0.099	0.099	0.099	0.099	0.098	0.135	0.135	0.130	0.131	0.131	0.134	0.123	0.123	0.122
Rajasthan	0.088	0.088	0.088	0.077	0.077	0.079	0.087	0.087	0.080	0.085	0.085	0.093	0.091	0.091	0.096
Tamil Nadu	0.141	0.141	0.141	0.118	0.118	0.109	0.120	0.120	0.108	0.102	0.102	0.105	0.133	0.133	0.129
Uttar Pradesh	0.113	0.113	0.113	0.109	0.109	0.097	0.115	0.115	0.103	0.097	0.097	0.093	0.118	0.118	0.099
West Bengal	0.089	0.089	0.089	0.082	0.082	0.075	0.126	0.126	0.108	0.089	0.089	0.081	0.098	0.098	0.085
Urban															
Andhra Pradesh	0.156	0.156	0.156	0.158	0.158	0.153	0.224	0.224	0.200	0.208	0.208	0.204	0.169	0.169	0.156
Assam	0.135	0.135	0.135	0.160	0.160	0.165	0.160	0.160	0.148	0.186	0.186	0.182	0.218	0.218	0.210
Bihar	0.137	0.137	0.137	0.158	0.158	0.154	0.177	0.177	0.154	0.184	0.184	0.163	0.137	0.137	0.120
Chhattisgarh	0.131	0.131	0.131	0.148	0.148	0.148	0.246	0.246	0.228	0.146	0.146	0.140	0.272	0.272	0.247
Gujarat	0.128	0.128	0.128	0.149	0.149	0.143	0.171	0.171	0.150	0.167	0.167	0.152	0.151	0.151	0.121
Haryana	0.123	0.123	0.123	0.145	0.145	0.141	0.207	0.207	0.198	0.230	0.230	0.234	0.271	0.271	0.265
Jharkhand	0.169	0.169	0.169	0.213	0.213	0.214	0.194	0.194	0.191	0.202	0.202	0.195	0.209	0.209	0.194
Karnataka	0.157	0.157	0.157	0.171	0.171	0.172	0.234	0.234	0.218	0.231	0.231	0.219	0.288	0.288	0.273
Kerala	0.167	0.167	0.167	0.171	0.171	0.173	0.267	0.267	0.251	0.244	0.244	0.234	0.269	0.269	0.247
Madhya Pradesh	0.166	0.166	0.166	0.166	0.166	0.164	0.233	0.233	0.216	0.246	0.246	0.251	0.263	0.263	0.280
Maharashtra	0.190	0.190	0.190	0.206	0.206	0.211	0.236	0.236	0.226	0.254	0.254	0.253	0.237	0.237	0.233
Odisha	0.142	0.142	0.142	0.156	0.156	0.156	0.198	0.198	0.187	0.268	0.268	0.274	0.210	0.210	0.211
Punjab	0.115	0.115	0.115	0.153	0.153	0.150	0.196	0.196	0.180	0.230	0.230	0.223	0.169	0.169	0.157
Rajasthan	0.131	0.131	0.131	0.136	0.136	0.130	0.186	0.186	0.165	0.176	0.176	0.161	0.179	0.179	0.154
Tamil Nadu	0.176	0.176	0.176	0.167	0.167	0.165	0.217	0.217	0.202	0.166	0.166	0.159	0.173	0.173	0.164
Uttar Pradesh	0.166	0.166	0.166	0.199	0.199	0.200	0.234	0.234	0.229	0.307	0.307	0.323	0.313	0.313	0.318
West Bengal	0.189	0.189	0.189	0.169	0.169	0.160	0.232	0.232	0.202	0.278	0.278	0.260	0.260	0.260	0.247

Note: "Cons" denotes the consumption index; "Exp-h' denotes the expenditure homothetic index, and "Exp-nh' denotes the expenditure non-homothetic index.

Table 5: Percentage changes in relative prices and inequality

Dep.var: %-changes in relative inequality	**Combined**	**Rural**	**Urban**
	(1)	(2)	(3)
%-changes in relative prices	0.259	0.231	0.259
	(0.045)	(0.029)	(0.030)
Constant	1.003	0.995	1.000
	(0.008)	(0.005)	(0.005)
R^2	0.396	0.545	0.593
N	51	51	51

Note: "Relative prices" shows the percentage changes in the price ratio of cereals over miscellaneous non-foods, whereas "relative inequality" shows the percentage changes in the ratio of the non-homothetic Theil index over the homothetic Theil index. All numbers are population weighted, using the multipliers provided by the NSS.

Table 6: Population shares and expenditure ratios

	1993–94	**1999–00**	**2004–05**	**2009–10**	**2011–12**
	(1)	(2)	(3)	(4)	(5)
Rural households in crop production					
Population share					
of rural population	0.68	0.62	0.55	0.53	0.44
of total population	0.51	0.47	0.42	0.39	0.31
Per capita expenditure ratio					
of rural non-crops	0.90	0.89	0.87	0.90	0.90
of all non-crops	0.77	0.72	0.69	0.69	0.71
Rural crop labours					
Population share					
of rural population	0.30	0.29	0.22	0.23	0.16
of total population	0.23	0.22	0.17	0.17	0.11
Per capita expenditure ratio					
of rural non-crops	0.77	0.78	0.72	0.76	0.77
of all non-crops	0.69	0.64	0.59	0.59	0.60

Note: The expenditure ratios show the ratio of the average per capita expenditure of rural crop producers over the average per capita expenditure of households outside crop production. All numbers are population weighted, using the multipliers provided by the NSS.

Table 7: Expenditure ratios vs. relative prices

	Engaged in crops		**Crops labours**	
Dep. var.: %-changes in expenditure ratios	Rural	All	Rural	All
	(1)	(2)	(3)	(4)
%-changes in relative prices	0.089	0.198	0.243	0.348
	(0.056)	(0.068)	(0.064)	(0.077)
Constant	0.004	-0.030	0.004	-0.038
	(0.010)	(0.010)	(0.011)	(0.011)
Observations	51	51	51	51
R^2	0.050	0.149	0.229	0.293

Note: The table shows the results from regressing the percentage change in the expenditure ratio (ratio of the average per capita expenditure of rural crop producers over the average per capita expenditure of households outside crop production) on the percentage change in relative prices (cereals over miscellaneous non-food). Robust standard errors are shown in parentheses.

Table 8: Theil index vs. relative prices

	Exp-h		**Exp-nh**	
Dep. var.: %-changes in Theil index	Rural	All	Rural	All
	(1)	(2)	(3)	(4)
%-changes in relative prices	-0.378	-0.480	-0.126	-0.200
	(0.168)	(0.132)	(0.171)	(0.131)
Constant	0.030	0.110	0.029	0.106
	(0.030)	(0.022)	(0.030)	(0.022)
Observations	51	51	51	51
R^2	0.094	0.214	0.011	0.045

Note: The table shows the results from regressing the percentage change in the Theil index on the percentage change in relative prices (cereals over miscellaneous non-food). Robust standard errors are shown in parentheses. "Exp-h" and "Exp-nh" denote estimates from the expenditure homothetic and expenditure non-homothetic indices, respectively.

Table 9: Cereal budget share regression

Dep. var: Cereal budget share	(1)	(2)
Crops labours	0.0595	0.0152
	(0.0004)	(0.0003)
Crops self-employed	0.0286	0.0143
	(0.0003)	(0.0003)
Log per capita expenditure		-0.3693
		(0.0015)
Log per capita expenditure squared		0.0204
		(0.0001)
Constant	0.0987	1.7195
	(0.0010)	(0.0053)
N	450.211	450.210
R^2	0.447	0.677
Unit fixed effects	yes	yes

Note: Robust standard errors are shown in parentheses.

Table 10: Unit values: medians, "Deaton et al. (2004)-adjustment" and our iteration procedure

		Rural					Urban				
	$(\hat{\beta})$	1993-94	1999-00	2004-05	2009-10	2011-12	1993-94	1999-00	2004-05	2009-10	2011-12
Rice											
UV Median		6.9	11.1	10.9	18.4	20.7	7.7	12.4	12.1	21.5	23.7
$\hat{p}_1$ (Deaton et al. (2004))	(0.191)	8.7	12.5	12.1	18.5	19.2	9.3	13.5	13.0	20.2	21.0
$\hat{p}_2$	(0.194)	7.6	11.8	11.7	19.1	20.4	8.2	12.9	12.8	21.3	22.6
$\hat{p}_3$	(0.194)	7.4	11.7	11.6	19.3	20.6	8.1	12.8	12.8	21.5	22.9
$\hat{p}_4$	(0.194)	7.4	11.7	11.6	19.3	20.6	8.1	12.8	12.8	21.5	22.9
$\hat{p}_5$	(0.194)	7.4	11.7	11.6	19.3	20.6	8.1	12.8	12.8	21.5	22.9
$\hat{p}_6$	(0.194)	7.4	11.7	11.6	19.3	20.6	8.1	12.8	12.8	21.5	22.9
Wheat											
UV Median		5.1	9.2	9.8	16.1	17.0	5.7	9.9	10.5	17.2	18.3
$\hat{p}_1$ (Deaton et al. (2004))	(0.072)	5.4	9.2	9.8	15.9	16.2	5.8	9.9	10.5	16.8	17.5
$\hat{p}_2$	(0.073)	5.1	8.9	9.7	16.1	16.6	5.5	9.7	10.5	17.1	18.0
$\hat{p}_3$	(0.073)	5.1	8.9	9.7	16.2	16.6	5.5	9.7	10.5	17.1	18.1
$\hat{p}_4$	(0.073)	5.1	8.9	9.7	16.2	16.6	5.5	9.7	10.5	17.2	18.1
$\hat{p}_5$	(0.073)	5.1	8.9	9.7	16.2	16.6	5.5	9.7	10.5	17.2	18.1
$\hat{p}_6$	(0.073)	5.1	8.9	9.7	16.2	16.6	5.5	9.7	10.5	17.2	18.1
Milk											
UV Median		6.6	10.6	11.6	19.0	25.1	8.0	12.3	13.7	20.9	27.0
$\hat{p}_1$ (Deaton et al. (2004))	(0.074)	7.1	11.0	12.4	18.5	23.7	8.3	12.3	13.9	20.0	25.5
$\hat{p}_2$	(0.075)	6.8	10.7	12.2	18.7	24.2	7.9	12.1	13.8	20.5	26.3
$\hat{p}_3$	(0.075)	6.7	10.7	12.2	18.8	24.3	7.9	12.1	13.8	20.5	26.4
$\hat{p}_4$	(0.075)	6.7	10.7	12.2	18.8	24.3	7.9	12.1	13.8	20.5	26.4
$\hat{p}_5$	(0.075)	6.7	10.7	12.2	18.8	24.3	7.9	12.1	13.8	20.5	26.4
$\hat{p}_6$	(0.075)	6.7	10.7	12.2	18.8	24.3	7.9	12.1	13.8	20.5	26.4
Fish, prawn											
UV Median		24.2	38.8	47.2	75.5	95.6	29.2	40.6	47.9	80.2	101.6
$\hat{p}_1$ (Deaton et al. (2004))	(0.191)	29.4	42.5	49.5	72.3	89.2	33.4	43.5	50.6	75.7	90.5
$\hat{p}_2$	(0.194)	25.7	40.0	48.0	74.8	94.7	29.7	41.5	49.8	79.7	97.6
$\hat{p}_3$	(0.194)	25.3	39.7	47.9	75.3	95.5	29.3	41.3	49.8	80.4	98.8
$\hat{p}_4$	(0.193)	25.3	39.7	47.9	75.3	95.6	29.2	41.3	49.8	80.5	98.9
$\hat{p}_5$	(0.193)	25.3	39.7	47.9	75.3	95.6	29.2	41.3	49.8	80.5	98.9
$\hat{p}_6$	(0.193)	25.3	39.7	47.9	75.3	95.6	29.2	41.3	49.8	80.5	99.0
Mustard oil											
UV Median		32.4	41.4	56.4	65.4	84.0	33.2	43.1	60.6	68.7	82.9
$\hat{p}_1$ (Deaton et al. (2004))	(0.004)	32.1	40.9	55.7	61.3	85.0	31.4	42.2	60.7	69.4	83.8
$\hat{p}_2$	(0.004)	32.0	40.8	55.7	61.3	85.1	31.3	42.1	60.7	69.5	84.0
$\hat{p}_3$	(0.004)	32.0	40.8	55.7	61.4	85.2	31.3	42.1	60.7	69.5	84.0
$\hat{p}_4$	(0.004)	32.0	40.8	55.7	61.4	85.2	31.3	42.1	60.7	69.5	84.0
$\hat{p}_5$	(0.004)	32.0	40.8	55.7	61.4	85.2	31.3	42.1	60.7	69.5	84.0
$\hat{p}_6$	(0.004)	32.0	40.8	55.7	61.4	85.2	31.3	42.1	60.7	69.5	84.0
Sugar											
UV Median		12.9	16.6	18.8	34.8	32.8	12.8	16.5	18.6	34.5	32.8
$\hat{p}_1$ (Deaton et al. (2004))	(0.003)	13.2	16.5	18.6	34.7	32.8	13.1	16.4	18.7	34.5	32.7
$\hat{p}_2$	(0.003)	13.1	16.5	18.6	34.8	32.9	13.1	16.4	18.6	34.5	32.8
$\hat{p}_3$	(0.003)	13.1	16.5	18.6	34.8	32.9	13.1	16.4	18.6	34.5	32.8
$\hat{p}_4$	(0.003)	13.1	16.5	18.6	34.8	32.9	13.1	16.4	18.6	34.5	32.8
$\hat{p}_5$	(0.003)	13.1	16.5	18.6	34.8	32.9	13.1	16.4	18.6	34.5	32.8
$\hat{p}_6$	(0.003)	13.1	16.5	18.6	34.8	32.9	13.1	16.4	18.6	34.5	32.8
Firewood and chips											
UV Median		0.6	1.0	1.1	2.0	2.7	0.9	1.3	1.4	2.5	3.4
$\hat{p}_1$ (Deaton et al. (2004))	(0.097)	0.7	1.1	1.3	2.0	2.6	0.9	1.4	1.7	2.4	3.2
$\hat{p}_2$	(0.099)	0.7	1.0	1.3	2.0	2.7	0.9	1.3	1.7	2.5	3.3
$\hat{p}_3$	(0.099)	0.7	1.0	1.3	2.0	2.7	0.9	1.3	1.7	2.5	3.3
$\hat{p}_4$	(0.099)	0.7	1.0	1.3	2.0	2.7	0.9	1.3	1.7	2.5	3.4
$\hat{p}_5$	(0.099)	0.7	1.0	1.3	2.0	2.7	0.9	1.3	1.7	2.5	3.4
$\hat{p}_6$	(0.099)	0.7	1.0	1.3	2.0	2.7	0.9	1.3	1.7	2.5	3.4
Ready-made garments											
UV Median		49.4	91.9	108.2	181.1	183.0	67.7	127.0	137.5	234.8	229.4
$\hat{p}_1$ (Deaton et al. (2004))	(0.473)	87.0	124.0	136.3	179.9	152.7	98.6	140.9	146.1	195.1	167.0
$\hat{p}_2$	(0.480)	62.3	106.6	125.7	196.1	176.5	73.2	125.3	140.5	222.1	201.8
$\hat{p}_3$	(0.479)	60.1	105.2	125.0	199.0	180.5	70.9	124.0	140.5	226.6	207.7
$\hat{p}_4$	(0.479)	59.8	105.0	124.9	199.3	181.0	70.6	123.9	140.4	227.2	208.4
$\hat{p}_5$	(0.479)	59.8	105.0	124.9	199.3	181.0	70.6	123.9	140.4	227.3	208.5
$\hat{p}_6$	(0.479)	59.8	105.0	124.9	199.3	181.1	70.5	123.9	140.4	227.3	208.5

Note: The table shows different item price estimates. "UV Median" shows the average over median unit values within each unit, whereas $\hat{p}_i$ shows the price estimates from the ith iteration of the procedure explained in Section 5.2. $\hat{b}$ present the quality-expenditure elasticity from each of these iterations.

Table 11: Number of households

	1993–94	**1999–00**	**2004–05**	**2009–10**	**2011–12**
	(1)	(2)	(3)	(4)	(5)
Total	97.965	100.954	99.788	80.386	80.409
1 adult & 0 children	6.166	6.052	5.620	4.683	4.613
2 adults & 0 children	8.432	7.887	8.567	7.988	8.078
2 adults & 1 child	6.329	6.412	6.300	5.339	5.614
2 adults & 2 children	9.015	9.971	10.182	8.726	8.695
2 adults & 3 children	6.851	7.212	6.684	4.492	4.189
3 adults & 0 children	5.259	5.162	5.626	5.451	5.814
3 adults & 1 child	4.075	4.458	4.447	3.915	4.173
3 adults & 2 children	4.458	4.537	4.688	3.857	3.647
4 adults & 0 children	4.656	4.762	5.373	5.270	5.439
4 adults & 1 child	3.708	3.966	3.952	3.769	3.934
4 adults & 2 children	3.407	3.321	3.694	3.181	3.149

Note: The table shows the number of households within each of the subsamples.

Table 12: PDS statistics

	Share of HHs		**Avg pc q PDS**		**UV market**		**UV PDS**		**PDS HHs w market**	
	Rice	Wheat	Rice	Wheat	Rice	Wheat	Rice	Wheat	Rice	Wheat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Rural										
1993-94	0.21	0.01	3	2	7	4	5	5	0.76	0.06
1999-00	0.29	0.16	3	1	11	9	5	4	0.86	0.44
2004-05	0.22	0.11	4	3	11	9	6	5	0.74	0.32
2009-10	0.37	0.28	4	2	18	15	5	6	0.80	0.49
2011-12	0.43	0.34	4	2	20	16	6	7	0.81	0.54
Urban										
1993-94	0.25	0.01	3	2	8	5	5	5	0.80	0.09
1999-00	0.21	0.16	3	2	12	10	7	6	0.88	0.42
2004-05	0.14	0.07	4	2	12	11	6	5	0.80	0.39
2009-10	0.22	0.20	4	2	23	17	4	6	0.87	0.49
2011-12	0.25	0.22	3	2	24	19	6	7	0.85	0.52

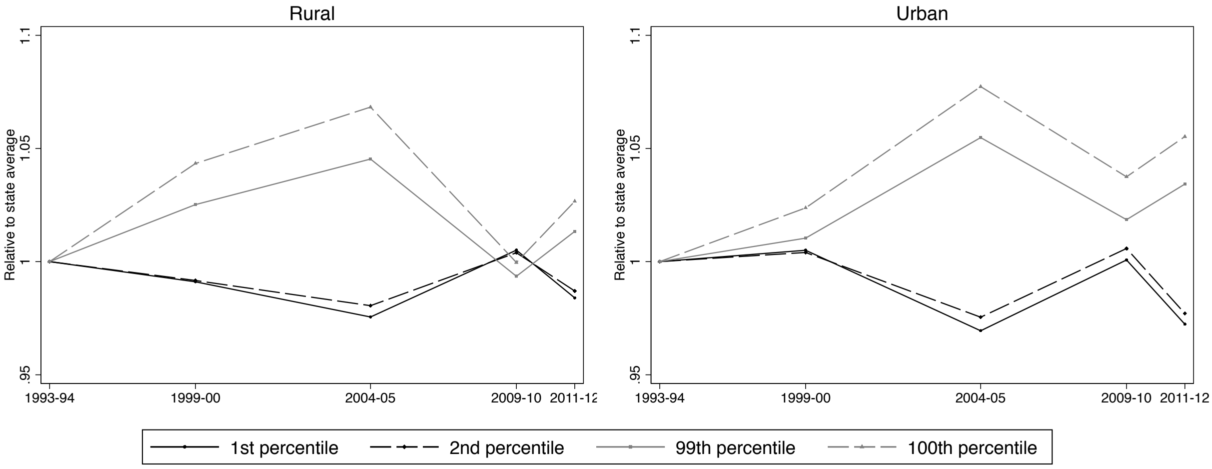
Note: "Share of HHs" displays the share of all households with any consumption of PDS rice and wheat, respectively. "Avg pc q PDS" presents the average per capita quantities (kg) for households with any PDS consumption. "UV market" and "UV PDS" show the average state and sector-specific median unit value for market purchases and PDS purchases, respectively. Finally, "PDS HHs w market" shows the fraction of households with any PDS consumption that report purchases of the same item in the market.

Table 13: Unit values and CPI price estimates

	$\mathbf{UV_{food}}$	$\mathbf{CPI_{food}}$	$\mathbf{CPI_{m.n\text{-}f}}$	$\frac{\mathbf{CPI_{food}}}{\mathbf{CPI_{m.n\text{-}f}}}$
	(1)	(2)	(3)	(4)
Rural				
1993–94	1.00	1.00	1.00	1.00
1999–00	1.56	1.51	1.65	1.10
2004–05	1.72	1.60	1.95	1.22
2009–10	2.91	2.61	2.64	1.01
2011–12	3.28	2.90	3.09	1.06
Urban				
1993–94	1.00	1.00	1.00	1.00
1999–00	1.55	1.61	1.69	1.05
2004–05	1.69	1.80	2.22	1.24
2009–10	2.88	2.95	3.21	1.09
2011–12	3.30	3.36	3.78	1.13

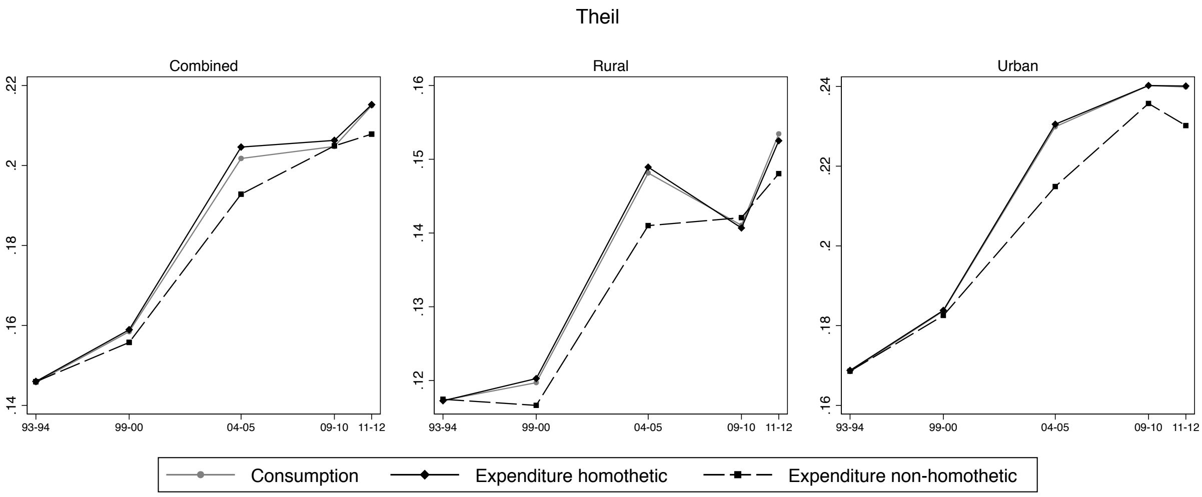
Note: "UV_{food}" presents the unit value food index, "CPI_{food}" presents the CPI food index, "$\text{CPI}_{m.n-f}$" presents the CPI sub-index that corresponds to our rest group, whereas $\frac{\text{CPI}_{food}}{\text{CPI}_{m.n\text{-}f}}$ displays the ratio of the CPI food index over the CPI rest group index.

Figure 1: Relative increases in cost-of-living

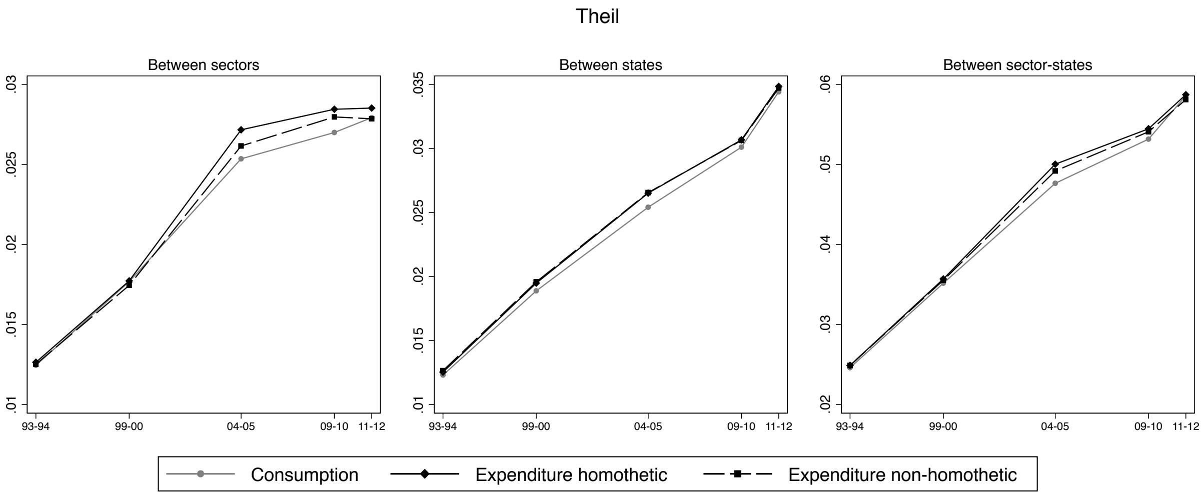


Note: The graphs show the relative increase in cost-of-living for some selected expenditure groups, relative to the average of all expenditure groups.

Figure 2: Trends in consumption inequality

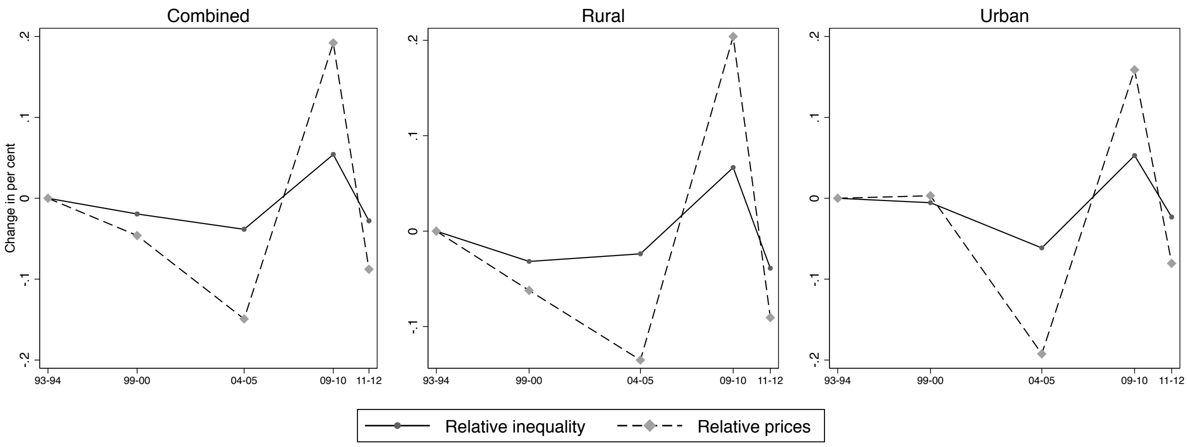


Note: The graphs show trends in real consumption inequality using the different cost-of-living indices. The left panel presents inequality for the rural and the urban sector combined, whereas the middle and the right panel display inequality separately for the two sectors.



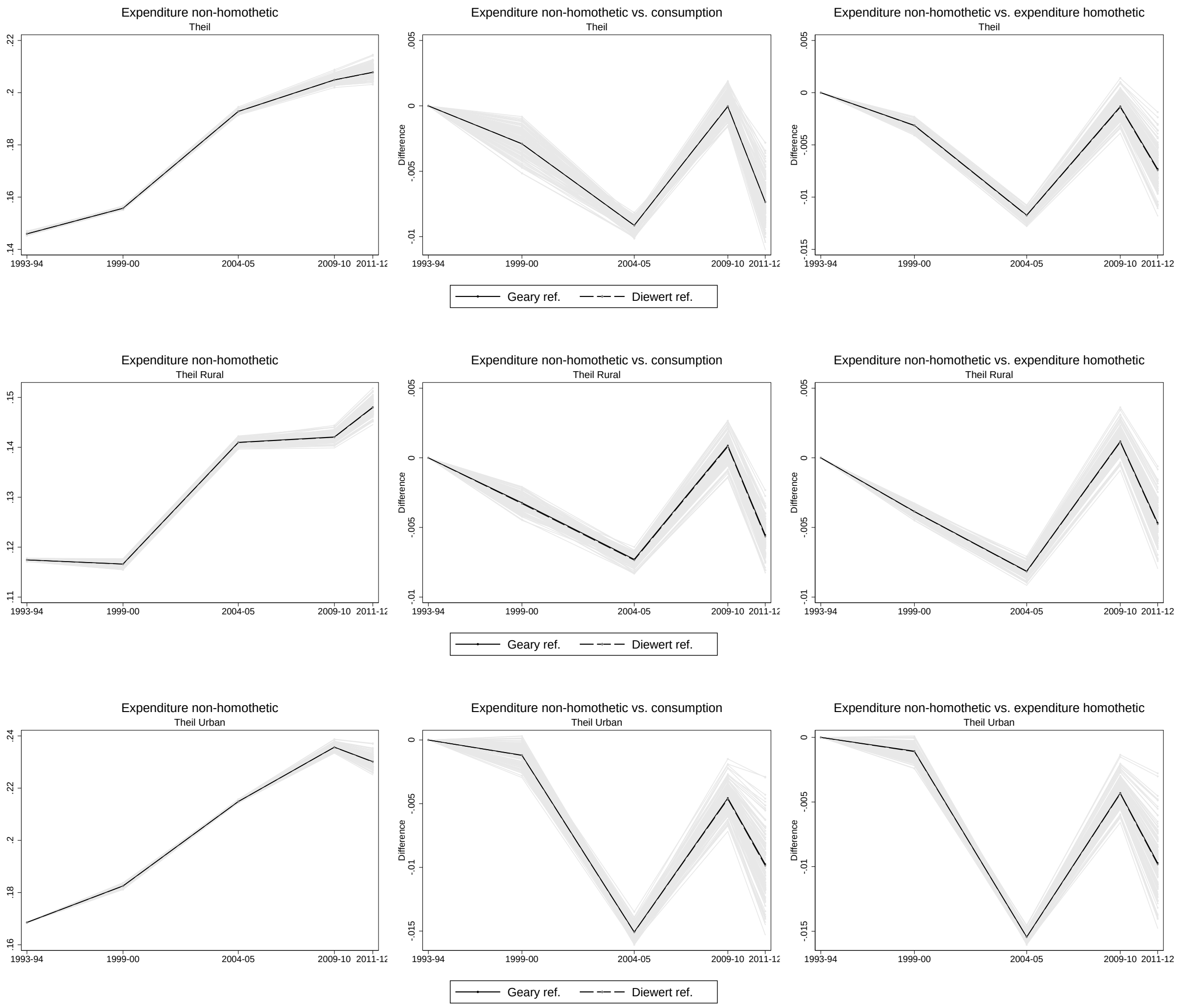
Note: The graphs show trends in between-group inequality using the different cost-of-living indices. “Sectors” shows inequality in average real consumption between the rural and the urban sector; “States” shows inequality in averages between states (rural and urban sector combined); whereas “Sector-states” presents inequality between every state and sector (what we call “units”).

Figure 3: Percentage changes in relative prices and inequality



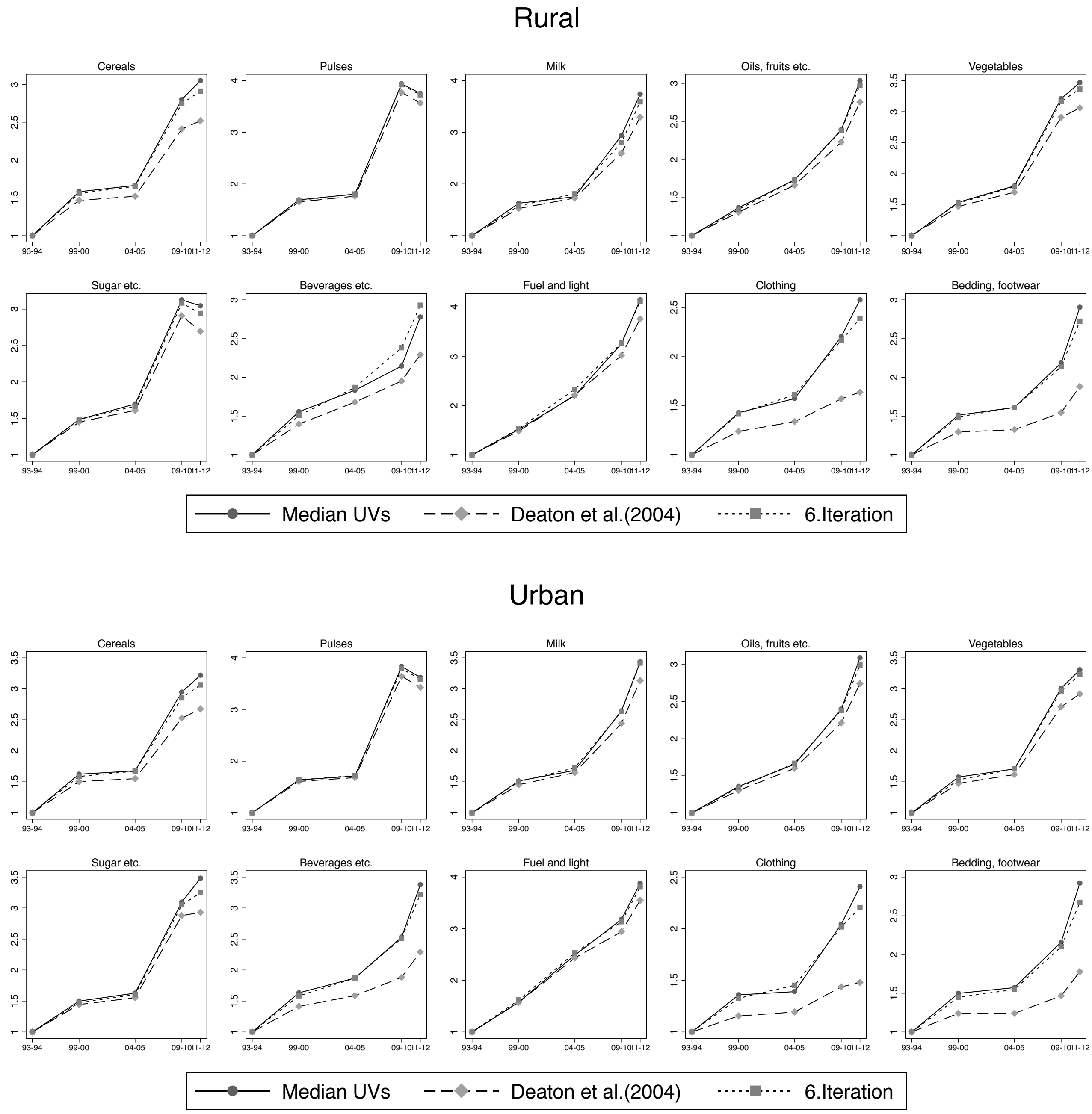
Note: "Relative prices" shows the percentage changes in the price ratio of cereals over miscellaneous non-foods, whereas "relative inequality" shows the percentage changes in the ratio of the non-homothetic Theil index over the homothetic Theil index. All numbers are population weighted, using the multipliers provided by the NSS.

Figure 4: Theil index, reference prices



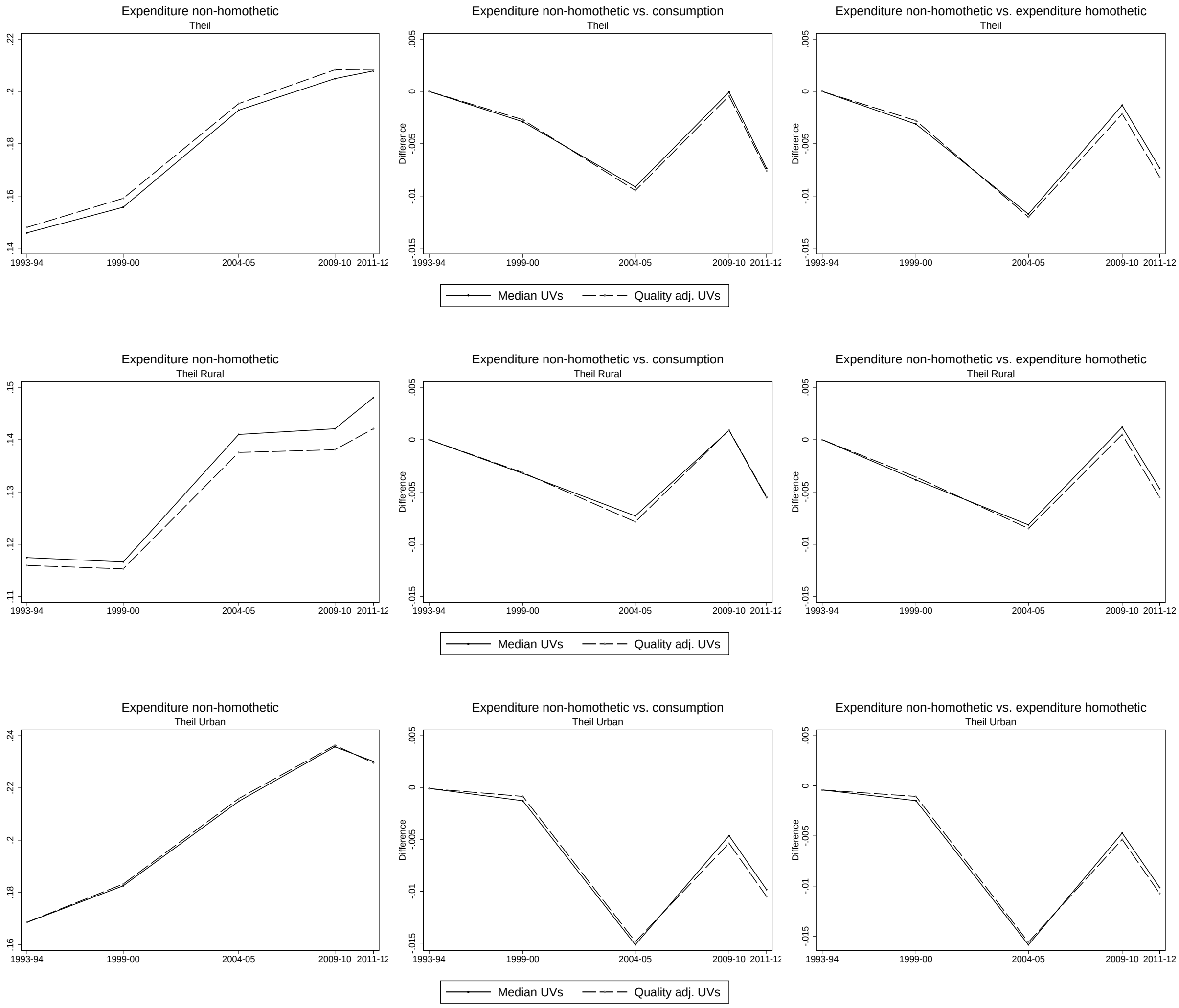
Note: The left panel in the graphs shows trends in inequality using the different reference price vectors and the expenditure non-homothetic cost-of-living index. The middle panel shows the differences between these estimates and those derived through the consumption index, whereas the right panel presents the differences versus the estimates derived through the expenditure homothetic index.

FIGURE 5: Trends in consumption group prices



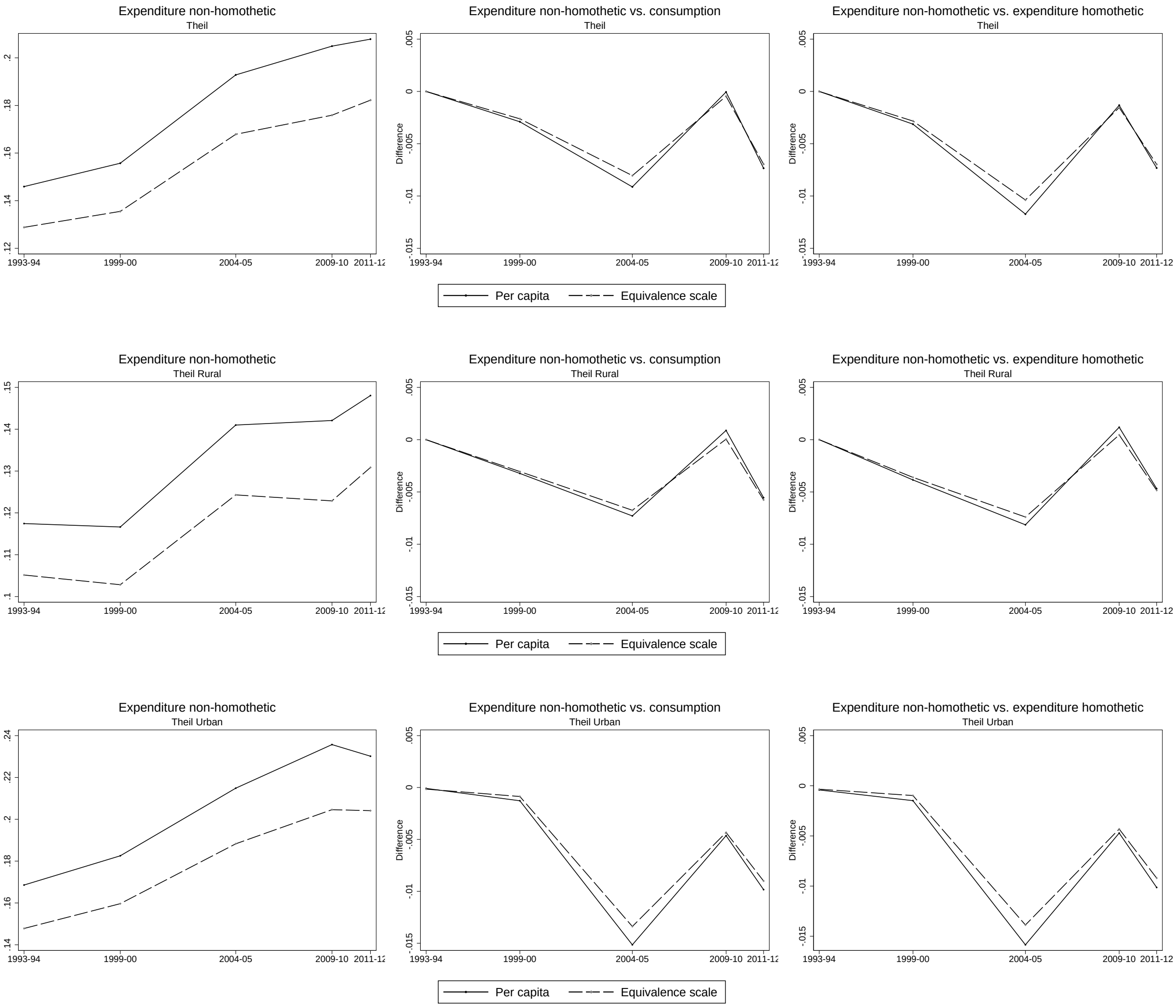
Note: The graphs show price trends for the ten first consumption groups. "Median UVs" shows trends using median unit values within each unit, "Deaton et al. (2004)" shows trends using the quality adjustment suggested by Deaton and co-authors, whereas "6.Iteration" displays the price trends when using the price estimates from the 6th iteration in our proposed procedure ($\hat{p}_6$).

Figure 6: Theil index, quality-adjusted unit values



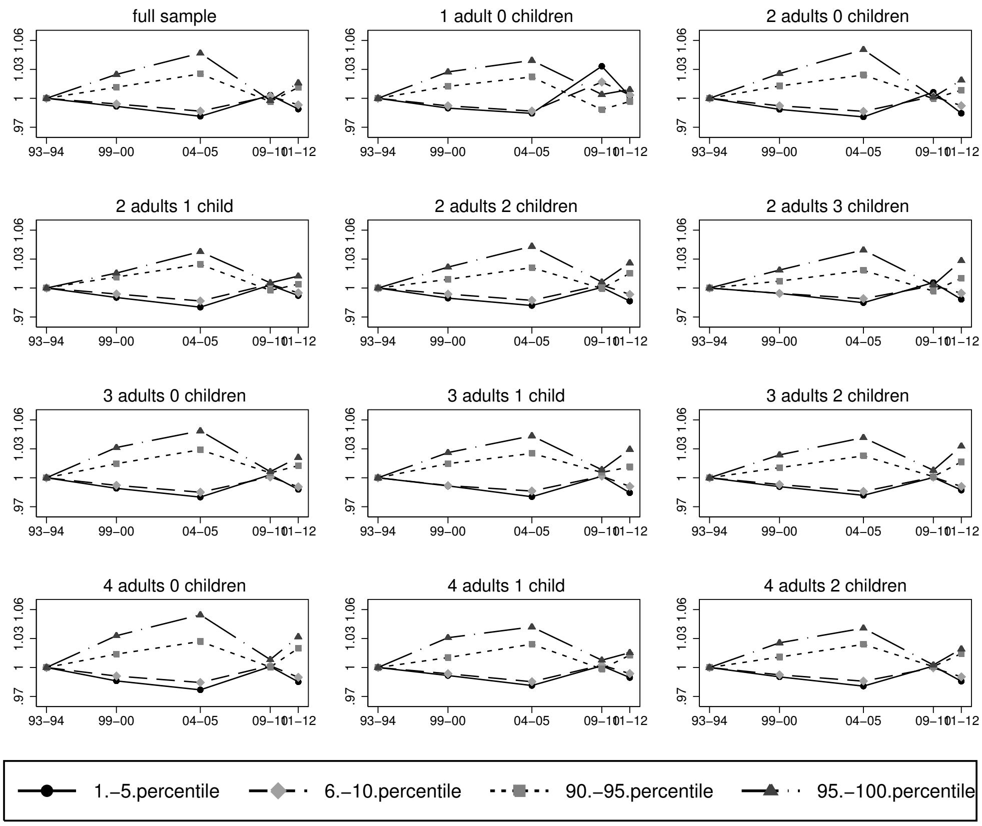
Note: The left panel in the graphs shows trends in inequality using the expenditure non-homothetic cost-of-living index, based on the median unit values and the quality-adjusted unit values. The middle panel shows the differences between these estimates and those derived through the consumption index, whereas the right panel presents the differences versus the inequality estimates derived through the expenditure homothetic index.

FIGURE 7: Theil index, equivalence scaling



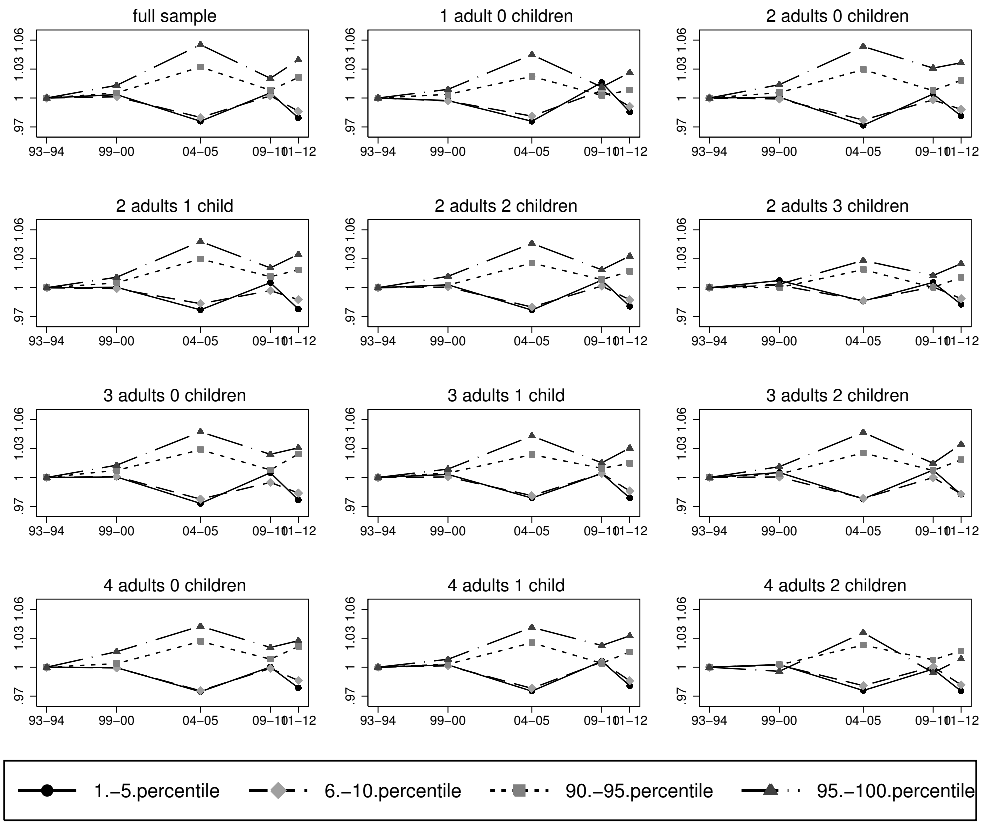
Note: The left panel in the graphs shows trends in inequality using the expenditure non-homothetic cost-of-living index, based on per capita expenditure and equivalence scaled expenditure. The middle panel shows the differences between these estimates and those derived through the consumption index, whereas the right panel presents the differences versus the inequality estimates derived through the expenditure homothetic index.

FIGURE 8: Relative increases in cost-of-living, Rural



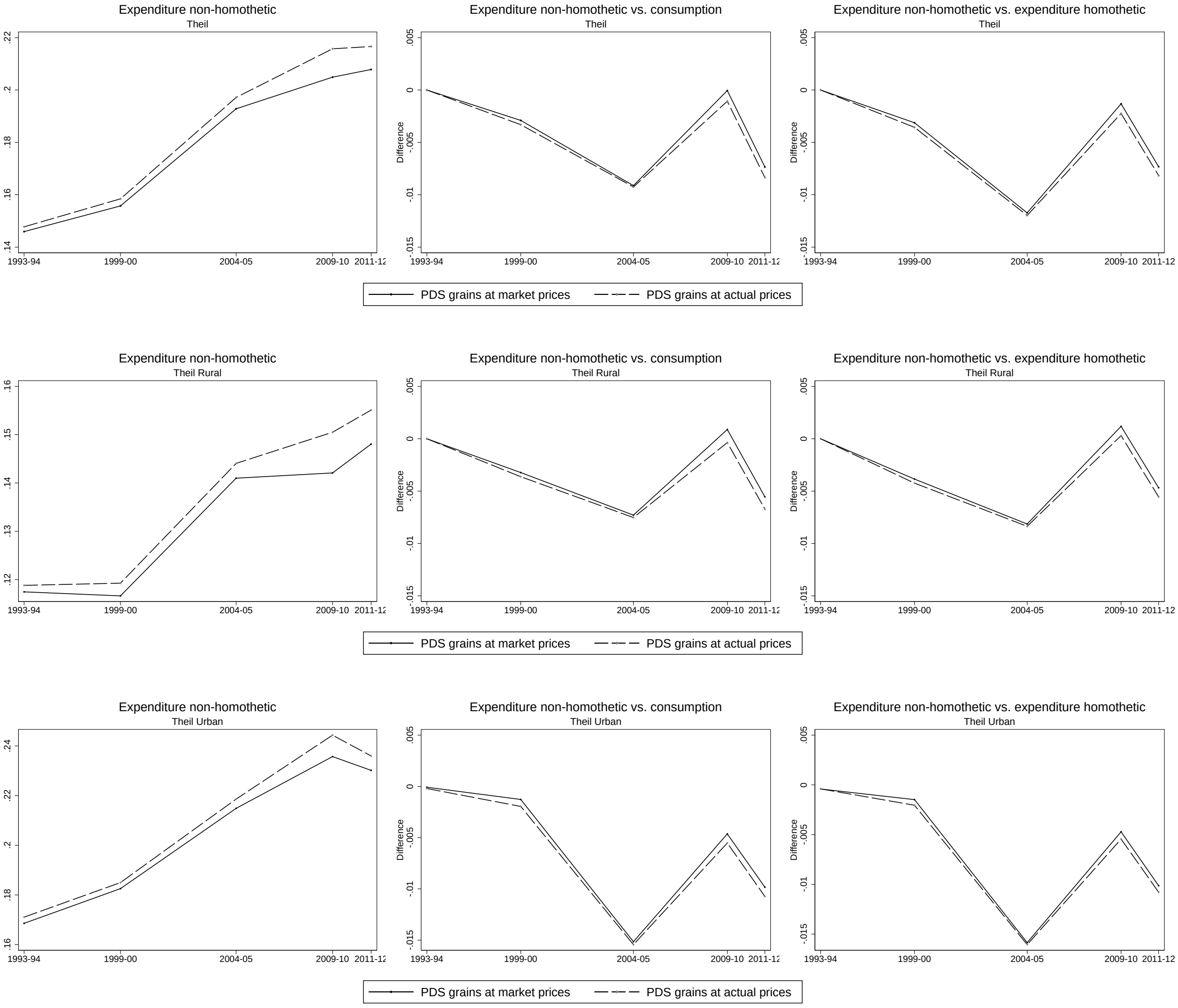
Note: The graphs show the relative increase in cost-of-living for some selected expenditure groups, relative to the average of all expenditure groups.

Figure 9: Relative increases in cost-of-living, Urban



Note: The graphs show the relative increase in cost-of-living for some selected expenditure groups, relative to the average of all expenditure groups.

FIGURE 10: Theil index, the PDS



Note: The left panel in the graphs shows trends in inequality using the expenditure non-homothetic cost-of-living index, based on the PDS valued at market prices and at actual prices paid. The middle panel shows the differences between these estimates and those derived through the consumption index, whereas the right panel presents the differences versus the inequality estimates derived through the expenditure homothetic index.