A New Engel on Price Index and Welfare Estimation[*]

David Atkin[†], Benjamin Faber[‡], Thibault Fally[§] and Marco Gonzalez-Navarro[¶]

March 2020

Abstract

Measuring changes in welfare, and particularly the price index, is challenging in the absence of well-measured prices covering the entire consumption basket. In this paper, we propose and implement a new approach that uses rich, but widely available, expenditure survey microdata to estimate theory-consistent changes in income-group specific price indices and welfare. We build on existing work that uses linear Engel curves and changes in expenditure on income-elastic goods to infer unobserved real incomes. A major shortcoming of this approach is that, while based on non-homothetic preferences, the price indices it recovers are homothetic and hence are neither theory consistent nor suitable for distributional analysis. In contrast, we show that we can recover income-specific price index and welfare changes from horizontal shifts in Engel curves if preferences are quasi-separable (Gorman, 1970; 1976) and we focus on what we term "relative Engel curves". Our approach is flexible enough to allow for the non-linear Engel curves we document in the data, and for non-parametric estimation at each point of the income distribution. We implement this approach to estimate changes in cost of living and household welfare in rural India between 1987/1988 and 1999/2000, and to revisit the impacts of India's trade reforms. Our estimates reveal that lower inflation rates for the rich erased the convergence between rich and poor seen in nominal incomes over this period, a pattern missed by the existing literature that calculates inflation using only well-measured food and fuels products.

Keywords: Household welfare, price indices, non-homothetic preferences, Engel curves, gains from trade.
JEL Classification: F63, O12, E31, D12.

1 Introduction

Measuring changes in household welfare is valuable in many contexts, both to evaluate the impacts of policies and to assess changes in well-being across time and space. Furthermore, given recent political upheaval and a renewed focus on inequality, there is an increased urgency to capture not just average changes but the full distribution. Measuring changes in household real income, however, requires extremely detailed microdata that are seldom, if ever, available. In particular, although we often have reliable data on changes in nominal incomes, estimating changes in the denominator of real income—the cost of living—requires knowledge of price changes for every item in household expenditures down to the variety level. If we are interested in distributional analysis, such detail is paramount since we know that different income groups consume very different bundles. If we take seriously the fact that products change in quality, new products appear and old ones disappear, and new modes such as online shopping arrive, then even knowledge of all price changes is not sufficient.

One promising avenue is to utilize rich and newly available microdata on consumption prices and quantities.[1] While such data are available for some countries and for some components of household welfare—e.g. US retail consumption using scanner microdata covering roughly 15 percent of consumption, or developing-country expenditure surveys on well-measured basic foodstuffs and fuels covering over 50 percent of consumption—these types of data are infeasible to collect for the entire consumption basket. Accurately measuring prices and quantities for services is particularly fraught with difficulty. Furthermore, even in the richest data environments, evaluating changes in welfare from observed price data still requires strong functional form assumptions (e.g. quality-adjusting prices for manufactures like electronics or accounting for the gains from variety).

In this paper, we instead propose and implement a new approach that uses rich, but widely available, expenditure survey microdata—and in particular does not require observing reliable price data for all consumption categories[2]—to estimate theory-consistent changes in exact household price indices for the full consumption basket, as well as welfare, at every point of the income distribution. We then implement this approach to quantify changes in household welfare for Indian districts over time, and to revisit the impacts of India's 1991 trade reforms.

Our approach builds on a longstanding literature using Engel curves and expenditure changes on income-elastic goods—typically foodstuffs—to recover unobserved changes in real income (e.g. Hamilton, 2001; Costa, 2001). Hamilton's (2001) initial goal was to correct biases in the US consumer price index (CPI) arising from difficulties in measuring quality-adjusted prices in

---

[1]See related literature below for recent work in this space.

[2]As we discuss below, while our method recovers the full price index from expenditure data on a subset of consumption and total outlays alone, we also use price data from well-measured consumption categories to test our preference restrictions and identifying assumptions, and to compute correction terms if necessary.

consumption categories such as services and manufactures. This exercise has since been re-peated for many developed and developing countries (see references contained in Nakamura et al., 2016). Almås (2012) applies this approach to correct for unobserved biases in purchasing power parity comparisons across countries, while Young (2012) estimates real income growth in sub-Saharan Africa. Bils and Klenow (2001) apply a related approach using quality Engel curves.

The bulk of this literature estimates linear Engel curves generated by the Almost Ideal Demand System (AIDS).[3] While this approach leans heavily on non-homotheticity—if demand is homothetic Engel curves are horizontal and thus uninformative about changes in welfare—we show that existing applications only recover changes in the price index if those changes are uniform across households at different income levels. I.e. by assuming away the income-group specific price indices generated by non-homothetic demand.[4]

To make progress on these challenges, our analysis proceeds in four steps. In the first step, we document two motivating facts using Indian expenditure survey microdata. First, Engel curves are typically non-linear. Second, Engel curves shift over time within a given market, and across markets within the same period, and that those horizontal shifts are not uniform across households of different income levels.

In the second step, we propose a novel methodology that addresses the drawbacks of existing approaches and is consistent with the two motivating facts. In particular, we use observed horizontal shifts in Engel-like curves across time (or space) to recover theory-consistent changes in exact price indices and household welfare at each point of the income distribution.

To fix ideas, consider a "textbook" Engel curve—the relationship between budget shares (y-axis) and log nominal outlays per capita (x-axis)—for food at two different points in time in the same market. The horizontal distance between curves at any point in the income distribution reveals the change in log nominal outlays that holds the food share constant across the two sets of prices. First, in Lemma 1 we show that for any rational utility function, this horizontal distance is equal to the change in the price index at any point in the income distribution, but only under the assumption of constant relative prices across the two periods. However, if there are no relative price changes, shifts in Engel curves must be parallel—in violation of the second motivating fact—and changes in price indices must be uniform across the income distribution. Unfortunately, if relative prices are allowed to change, horizontal distances between textbook Engel curves do not in general recover changes in price indices (Lemma 2).

To make progress, we add additional structure to the very general preferences above in order to relax these restrictions on relative price changes. In our main propositions, we focus on

---

[3]Almås (2012) and Almås et al. (2018) use quadratic Engel curves and QUAIDS preferences.

[4]Almås et al. (2018) also note this shortcoming. Calculating non-uniform price index changes under the existing AIDS methodology re-introduces the need to observe the full vector of prices (see Section 2). Their paper addresses this challenge either by using what price information is available, bounding estimates or imposing additional structure on relative prices. Almås and Kjelsrud (2017) apply these approaches to measuring inequality in India.

the broad class of quasi-separable preferences (following Gorman 1970; 1976)[5] and what we term "relative Engel curves"—the relationship between expenditure shares within a given sub-set of goods or services G and log total nominal outlays per capita (i.e. the sum of spending on G and all other groups). If preferences are quasi-separable, we prove that as long as relative prices remain constant within group G, horizontal shifts in relative Engel curves of goods within G reveal changes in exact price indices (i.e. the price index for the full consumption basket). Critically, since preferences are non-homothetic and relative prices can vary arbitrarily outside of G, these horizontal shifts—and hence price indices—are income-group specific. It is then straightforward to recover changes in welfare for households at any point in the income distribution from the distance in outlays between period 0 and 1 relative expenditure shares, either traveling along period 0's relative Engel curve (to recover the equivalent variation, EV) or period 1's curve (to recover the compensating variation, CV).

Of course, relative prices are likely to be changing within group G. We show that if we average our estimated price index changes over many goods within the group, we can relax the assumption that relative prices are fixed within G with an orthogonality condition: that changes in relative prices are unrelated to the (local) slopes of relative Engel curves. This result provides us with the implicit identifying assumption required in data-poor environments where there is no reliable price data. However, if reliable price data are available for some subset of goods—such as food and fuels in the Indian setting or supermarket retail in the US setting—we can explicitly test this orthogonality assumption and correct for any violations. In particular, if we restrict the price index estimation to only product groups G where price changes are observed, the price data allow us to compute correction terms addressing any bias in the full price index coming from confounding within-G relative price changes.

Put another way, we obtain unbiased estimates of the full price index that covers *all* household consumption using only relative Engel curves and prices for subsets of goods for which we have reliable price data, without requiring any restrictions on relative price changes outside of G. We argue above that it is not possible to obtain reliable (i.e. quality- and variety-adjusted) price data for large swaths of the service and manufacturing industry. Thus, an alternative view of the results above is that we provide the minimal structure on preferences (i.e. quasi-separability) to allow us to uncover the full price index in the absence of prices for substantial subsets of consumption.

An obvious question to ask is how general is the quasi-separable class of preferences? Quasi-separability requires that subsets of goods or services are separable in the expenditure function (not the utility function), so that relative budget shares within a subset G of goods are functions of relative prices within G and household utility. This is less restrictive than the more common

---

[5]Deaton and Muellbauer (1980) also refer to quasi-separability as implicit separability. Blackorby et al. (1991) distinguish quasi and implicit separability. We describe this class of preferences in more detail below.

assumption of direct separability across goods in the utility function (hence the term "quasi"). The relative expenditures of goods within subset G are still a function of all prices in the rest of the economy but, crucially for our needs, prices outside the subset only affect relative expenditures within G through their effect on utility. Examples of preferences in this class are several variants of PIGL and PIGLOG (see Deaton and Muellbauer, 1980), non-homothetic CES preferences (e.g. Gorman, 1965; Hanoch, 1975; Comin et al., 2015), and a general class of Gorman preferences discussed in Fally (2018). Relative to these special cases, our approach is more general. We can, for example, allow for arbitrary own and cross-price effects within and outside of the subset G. And since quasi-separable demand systems can be arbitrarily non-linear—they can be of any rank in the terminology of Lewbel (1991)—, our approach is both flexible enough to allow for the non-linear Engel curves documented in the first motivating fact and amenable to non-parametric estimation of price indices at each point of the income distribution consistent with the second stylized fact.

In the third step, we form a bridge between the theoretical results and the empirical implementation. Our estimation approach follows directly from our main proposition and uses expenditure survey microdata to estimate non-parametric relative Engel curves for every period and every good inside product group G. The horizontal difference between curves across time (or space) at a particular point of the household income distribution reveals the change in the price index for those households. To take this approach to the data, we derive a set of testable requirements for unique and unbiased identification: i) on the invertibility of Engel curves, ii) on identification and correction terms when relative prices are changing within G, iii) on aggregating barcode-level data to goods-level data, iv) on testing quasi-separability, v) on sample selection when horizontal shifts can be identified only for a subset of goods, and vi) on preference heterogeneity across households and over time. As (ii) and (iv) require reliable price data for a subset of goods, we propose restricting estimation to goods where such data are available—but recall that the methodology still recovers the full price index for all consumption.

In the final step, we implement our methodology in two applications. First, we draw on Indian household expenditure survey microdata to quantify changes in rural welfare between 1987/88 and 1999/2000 at different points in the income distribution, and for every district in India. We compare our New Engel estimates to the leading existing Indian CPI estimates that come from Deaton (2003b) who calculates standard Paasche and Laspeyres price index numbers using changes in prices of products in the Indian household surveys with both reliable quantity information and no evidence of multiple varieties within a given market. For poorer deciles of the income distribution, we find very similar levels of consumer price inflation. Given that the products Deaton deems to have reliable prices—foods and fuels—cover about 80 percent of total outlays for poorer rural households in the sample, it is reassuring that our esti-

mates of the full price index for these households are very similar to Deaton's estimates of what is essentially a food and fuel price index (despite coming to this conclusion in very different ways—we exploit shifts in relative Engel curves while Deaton uses observed price changes).

Looking across the income distribution, our estimates bring to light that price inflation has been far from uniform, with significantly lower inflation rates for richer households—something that is not apparent from calculating standard price indices even when using income-group- and district-specific expenditure weights. Thus, while estimates based on standard price indices suggest that India saw significant convergence between poor and rich households over this period, we find that this convergence entirely disappears once we account for the differential inflation across income groups revealed by our approach. The most likely explanation for these findings is that higher-income Indian households disproportionately benefited from lower inflation in product categories such as services and manufactures where reliable price data are simply not available. This lower inflation is consistent with substantial increases in both the quality and variety of manufacturing products, and price declines, resulting from large reductions in tariff protection (see Goldberg et al., 2010); as well as rapid growth in the share of services in both GDP and employment over this period (Mukherjee, 2015). Standard approaches to price index estimation miss these patterns as these categories are either ignored entirely (as in Deaton, 2003b) or included without any quality or variety correction (as in India's official CPI). Since wealthy households spend disproportionately on these categories, difficulties in measuring service and manufacturing prices have the potential to change the distribution of welfare changes as we find.

As well as serving as a proof on concept for our methodology, this analysis sheds new light on the Great Indian Poverty Debate. Because India's 1999-2000 National Sample Survey (NSS) added an additional 7-day recall period for food products (which inflated answers to the consistently asked 30-day consumption questions), there has been much disagreement on how much poverty changed over the reform period (see Deaton and Kozel, 2005 for an overview of the debate). Given this recall issue, Deaton (2003a) calculates poverty by adjusting food expenditure using the mapping between food expenditure and fuels (for which no additional recall period was added) from earlier rounds. Such a method implicitly assumes that relative prices of food and fuels did not change.[6] In contrast, as long as the additional recall period did not change *relative* budget shares within a given food product group G, our approach remains unbiased. We show that this assumption holds by exploiting the fact that the 1998 'thin' survey round randomly assigned households to different recall periods. Thus, our approach has the additional benefit of dealing with the recall issues at the center of the Great Indian Poverty Debate.

In the second application, we use our New Engel method to revisit Topalova's (2010) analysis

---

[6]Tarozzi (2007) pursues a related approach that relies on mappings between poverty and multiple auxiliary variables being stable over time.

of the local labor market impacts of India's 1991 trade reforms. Topalova's main finding is that rural poverty rates (the fraction of households below the poverty line) increased relatively more in districts where import competition rose most. While Topalova highlights effects on poverty rates, our approach uncovers adverse effects of import competition across the full distribution of household income, including at the very top. Additionally, we find that the adverse effects on nominal expenditures are amplified by cost of living inflation. That is, areas adversely affected by import competition experienced higher local price inflation compared to less exposed areas. This somewhat surprising finding is also evident in Paasche and Laspeyres indices using food and fuel groups with well-measured prices selected by Deaton (2003b).

In addition to the literatures mentioned above, this paper relates to a large literature on the structure of demand and household preferences (e.g. Gorman, 1995; Blackorby et al., 1978; Lewbel and Pendakur, 2009), and provides several new results and proofs. Ligon (2019) explores an alternative approach based on Frisch demand that assumes preferences are directly separable and isoelastic in own prices to recover differences in the marginal utility of money ("neediness") between households. Jones and Klenow (2016) produce a consumption equivalent welfare measure that incorporates leisure, mortality and inequality. A recent literature uses barcode-level microdata for price index estimation (e.g. Argente and Lee, Forthcoming; Atkin et al., 2018; Jaravel, 2018). For example, Redding and Weinstein (2020) show how to use CES preferences to account for changes in product demand, quality, and variety when prices are observed. Our approach's ability to capture rich distributional effects also generates parallels with the literature on non-homothetic preferences and the gains from trade (e.g. Fajgelbaum and Khandelwal, (2016);[7] Borusyak and Jaravel, (2018); Hottman and Monarch, (2018)).

We structure the paper as follows. Section 2 provides a brief review of the existing Engel approach to price index estimation. Section 3 describes the data and presents stylized facts that motivate our theoretical approach. Section 4 develops the theory. Section 5 presents our estimation approach and derives corollaries for unique and unbiased identification. Section 6 applies our methodology in the two applications described above. Section 7 concludes.

2   Review of the Existing Engel Approach

In order to clarify our contribution, we briefly recap existing approaches that use Engel relationships to uncover changes in real income (e.g. Nakamura, 1996; Costa, 2001; Hamilton, 2001; Almås, 2012; Nakamura et al., 2016). These papers estimate linear Engel curves using AIDS and use the recovered income elasticity to infer changes in real income from changes in budget shares on income elastic goods.

---

[7]Fajgelbaum and Khandelwal use AIDS and country-level expenditure shares to compute welfare changes across the income distribution. As noted above, this implicitly requires knowledge of all prices, which they indirectly obtain from gravity equations.

To be more precise, under AIDS, Engel curves take the following form:

$$\frac{x_{hi}}{y_h} = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(y_h/\Lambda(p)) \tag{1}$$

where y_h is total nominal outlays per capita for household h, $\frac{x_{hi}}{y_h}$ is the household's budget share spent on good i (as a share of total expenditure), $\sum_j \gamma_{ij} \log p_j$ are own and cross-price effects and Λ is a price aggregator (not a price index) defined by:

$$\log \Lambda(p) = \alpha_0 + \sum_i \alpha_i \log p_j + \sum_{i,j} \gamma_{ij} \log p_i \log p_j$$

with $\sum_j \gamma_{ij} = 0$ for all i. Hence the literature estimates time series regressions of the form:

$$\frac{x_{hit}}{y_{ht}} = \alpha_{it} + \sum_j \gamma_{ij} \log p_{jt} + \beta_i \log y_{ht} + \epsilon_{cit}$$

where the constant α_{it} is allowed to vary across time, but not the slope β_i. It is straightforward to see that changes in the intercept over time scaled by income elasticities reveal changes in the price aggregator:

$$\frac{d\alpha_{it}}{\beta_i} = -d \log \Lambda(p_t)$$

with $d\log(y_{ht}/\Lambda(p_t)) = d\log y_{ht} + \frac{d\alpha_{it}}{\beta_i}$. If the constant is allowed to vary by location, the method can correct for PPP bias across countries as in Almås (2012).

There are several drawbacks to this approach. While total expenditures divided by a price aggregator is an appealing measure of "real income", it is not a theory-consistent welfare metric since it does not correspond to welfare in this demand system. This is apparent from the fact that AIDS is non-homothetic and so, in general, changes in the price index will differ across income groups, yet the price aggregator $\Lambda(p_t)$ is homothetic. More precisely, under the preferences represented by the AIDS expenditure function, the proportional change in household welfare is not $d\log(y_{ht}/\Lambda(p_t))$, but:

$$d \log U_{ht} = d\frac{\log(y_{ht}/\Lambda(p_t))}{\prod_j p_{jt}^{\beta_j}} \tag{2}$$

Essentially there are two price aggregators under AIDS, $\Lambda(p_t)$ and $\prod_j p_{jt}^{\beta_j}$, which are combined using household-specific weights to generate income-group specific price indices.

Only in the special case where price realizations are such that $\prod_j (p_{jt})^{\beta_j}$ is unchanged over time is $d\log(y_{ht}/\Lambda(p_t))$ proportional to $d\log U_{ht}$. In this case, the true change in the price index is uniform across income groups. This homotheticity-like restriction is unsatisfactory for a method that entirely depends on Engel curves being non-homothetic, a point also noted by

7

Almås et al. (2018). It is also unsuitable for evaluating distributional effects where it is important to allow for the possibility of different inflation rates across the income distribution.

The second drawback is that in order to estimate the AIDS system above, in principle all prices are needed for the cross-price controls $\sum_i \alpha_i \log p_j$. As we argue in the introduction, reliable price data for services and manufactures are rarely, if ever, available. This is particularly problematic if a researcher wanted to calculate the true AIDS price index, since large parts of the second price aggregator $\prod_j (p_{jt})^{\beta_j}$ cannot be accurately calculated. Finally, AIDS imposes linear Engel curves. In the next section, we will show that both this restriction and the homotheticity of price indices are inconsistent with empirical evidence, before proposing a new approach that overcomes these shortcomings.

3 Data and Motivating Facts

3.1 Data

Following Topalova (2010) and the Great Indian Poverty Debate, we draw on rural households in two of India's "thick" NSS survey rounds covering 1987/88 (43rd round) and 1999/2000 (55th round). Each round provides us with detailed expenditure data on approximately 80,000 households residing in more than 400 Indian districts. Households are asked about their expenditures on 310 goods and services in each survey round. Examples include wheat, coconut, turmeric, washing soap and diesel. The sum of all expenditures provides our measure of total household outlays. Given limited saving in India this will closely approximate nominal income (and even more closely, permanent income). For readability, in what follows we use the word outlays interchangeably with income. The surveys also contain detailed household characteristics, district of residence, and survey weights that we use to make the sample nationally representative.

Deaton (2003a) and Deaton and Tarozzi (2005), carefully analyze these NSS expenditure surveys to identify a subset of 136 food and fuel products for which both quantities are recorded and prices (obtained from expenditures divided by quantities) are robust to concerns about unobserved product quality or variety.[8] These goods cover on average 75 percent of household consumption in our sample. As we discuss in Sections 5 and 6, these products will be particularly valuable for our estimation as this subset of goods with reliable price data will allow us to directly test the implicit preference restrictions and identifying assumptions, and compute correction terms if necessary.

Finally, it is important to note that in the 55th round, the surveys included a 7-day recall period for all food products (in addition to the standard 30-day recall period asked across

---

[8]In particular, Deaton (2003a) and Deaton and Tarozzi (2005) discard product categories that are likely to contain multiple varieties or quality levels based on either the name of the category (e.g. "other milk products") or the observation of bi-modal distributions of prices within the category (e.g. "liquid petroleum").

rounds). While we only use the responses to the consistently-asked 30-day recall questions, Deaton (2003a,2003b) and others show that households inflated their 30-day reports of food expenditures to be consistent with their 7-day reports. This "recall bias" raises reported total nominal outlays (the numerator for evaluating changes in real incomes) even using the 30-day recall data, and is at the center of the Great Indian Poverty Debate. In Section 6, we show that our approach is robust to this recall bias as it relies on relative consumption patterns within product groupings that we find to be unaffected by the addition of the 7-day recall question.

3.2 Motivating Facts

In this subsection, we use the Indian NSS data to establish two facts that motivate our theoretical approach below. We start by plotting Engel curves following Working's (1943) now standard formulation where budget shares are on the y-axis and log total household outlays per capita are on the x-axis. Figure 1 shows non-parametric Engel curve estimates—kernel-weighted local polynomials—for salt, a good consumed widely across all parts of India and all income groups. As salt is an inferior good, Engel curves are downward-sloping with richer households spending a smaller share of their budget on the good. Panel A shows the Engel curves for the largest rural market, Midnapur, estimated separately for each survey round. Panel B shows the Engel curves in 1999/2000 for the largest market in the North, East, South and West of India. Two motivating facts are apparent from these figures.

Motivating Fact 1: Engel Curves Are Non-Linear

As Figure 1 shows, Engel curves can be non-linear. To test this observation more formally, we estimate the following regression for each of the 310 goods and services in the NSS surveys:

$$\frac{x_{hit}}{y_{ht}} = \theta_{imt} + \beta_{ki} F\left(\log\left(y_{ht}\right)\right) + \epsilon_{hit} \tag{3}$$

where x_{hit} is expenditure per capita on good i and y_{ht} is total nominal outlays per capita for household h residing in market m during survey round t.[9] θ_{imt} are good-by-market-by-period fixed effects (capturing variation in relative prices or tastes across markets) and $F(\cdot)$ is a vector of polynomial terms with the order indexed by k. Panel A in Table 1 presents the fraction of goods and services for which the data formally reject the null hypothesis that the second-order and above polynomial terms (up to $k = 4$) are jointly equal to zero—i.e. that Engel curves are linear). We can reject linearity at the 5 percent level of significance for 90 percent of goods and services. As noted previously by Banks et al. (1997) and others, Engel curves are frequently non-linear.

---

[9]Results are not sensitive to including a full vector of household characteristic controls on the right hand side of equation (3). We discuss potential taste heterogeneity across households in Section 5.2.

Motivating Fact 2: Shifts in Engel Curves Over Time Are Not Parallel

As is apparent from the plots for a single market across different time periods (Figure 1 Panel A), Engel curves shift over time, and they do not shift in parallel—i.e. the horizontal shift in Engel curves is not uniform across the income distribution. This fact is important because, as we discuss in detail in the theory, a household in the first period with the same budget share as a household in the second period will have the same level of utility, absent confounding relative price changes. Thus, the horizontal distance between curves is informative of the amount of money the second period household has to be given to be just as well off as the first period household—i.e. the change in the price index. If the shifts are not parallel, these price index changes potentially differ across the income distribution.

To provide a more formal test of whether shifts in Engel curves across time are parallel we flip the axes and run the following regression:

$$\log\left(y_{ht}\right) = \theta_{im} + \delta_{im}Post_t + \beta_{kim}F\left(\frac{x_{hit}}{y_{ht}}\right) + \gamma_{kim}Post_t \times F\left(\frac{x_{hit}}{y_{ht}}\right) + \epsilon_{hit}$$

where θ_{im} are market-by-good fixed effects, $Post_t$ is an indicator for the more recent survey round, and $F\left(\frac{x_{hit}}{y_{ht}}\right)$ is a vector of polynomial terms (up to the $k = 4$th order) of budget shares of good i. We test the hypothesis that the four γ_{kim} interactions between the budget share polynomials and $Post_t$ are jointly equal to zero (i.e. Engel curves only shift in parallel, as captured by δ_{im}). As reported in the second column in Table 1, we formally reject the null of a uniform shift for 69 percent of the market-by-good cells (at a 95 percent confidence level). Figure 1 of Panel B suggests that Engel curves do not shift in parallel across space, which has analogous implications for estimating price index differences across locations.

4   Theory

In this section we develop a new approach to estimating changes in price indices and welfare that i) addresses the drawbacks of the existing Engel approach discussed in Section 2, and ii) is consistent with the two motivating facts in Section 3. We proceed in three steps. First, we introduce in a very general setting—for any rational utility function—the logic for why horizontal shifts in Engel curves relate to changes in price indices and hence welfare. While appealing, we prove that such an approach uncovers theory-consistent price indices only under the restriction that all relative prices remain unchanged (Lemma 1), but not in general (Lemma 2)—a restriction that precludes income-group specific price index changes and necessarily violates Motivating Fact 2.

In the second step, we show how to relax this restriction and allow for income-group specific price index changes. To make progress, we focus on a broad class of quasi-separable prefer-

ences and on "relative Engel curves" that describe how relative expenditure shares within any subset of goods or services G vary with log total nominal outlays. Proposition 1 proves that if, and only if, preferences are quasi-separable, horizontal shifts in relative Engel curves within G recover exact price indices, as long as relative prices are fixed within G. By placing no restrictions on price realizations outside of G, price indices can vary arbitrarily with income and can be estimated at each point of the income distribution, accommodating the patterns documented in Motivating Fact 2. Moreover, quasi-separable preferences can be of any rank and so our approach is flexible enough to allow for highly non-linear Engel curves consistent with Motivating Fact 1. In the final step, Proposition 2 extends our results to account for relative price changes within G, which can be readily accommodated if high-quality price data are available for some subset of products (e.g. food or supermarket retail).

4.1 Using Shifts in Engel Curves to Infer Changes in Price Indices and Welfare

Consider comparing an Engel curve, for example food budget shares plotted against log total outlays, at two different points in time (or across space). The horizontal distance between curves at any point in the income distribution reveals the change in log total outlays which holds the food share constant. The close link between this distance and the price index is obvious in the case where there are no changes in relative prices. Then, as long as demand is homogeneous of degree zero in total outlays and prices, a uniform price increase is equivalent to an equally sized fall in outlays. Hence, between points in time (or across space) the price index change expressed in units of log outlays is exactly equal to the size of the horizontal shift. More generally, this will not be the case when relative prices are changing. This subsection develops these statements formally.

To match our empirical setting, we focus the discussion below on inferring price index changes over time for households at a given percentile of the income distribution within a particular market location.[10] Isomorphic results would hold across space if we replaced time periods by locations. In what follows, the subscript i indexes goods and services (for readability we will refer to them simply as goods), h indexes households, and superscripts 0 and 1 indicate time periods. We denote Engel curves as functions $\frac{x_{hi}}{y_h} = E_i(p, y_h)$ with budget shares on the y-axis and log outlays on the x-axis, where x_{hi} is household per-capita expenditure on good i, y_h is household nominal outlays per capita, and p is the full vector of consumption prices.

We define $P^1(p^0, p^1, y_h^1)$ (or in more concise notation just $P^1(y_h^1)$ or P^1) as the exact price index change between period 0 and period 1 prices holding utility at period 1's level (i.e. P^1 is defined implicitly by $V(p^1, y_h^1) = V(p^0, \frac{y_h^1}{P^1(y_h^1)})$ where V is the indirect utility function). In other words, the price index $P^1(y_h^1)$ converts the household's period 1 nominal income to the hypothetical level of income that would make them equally well off under period 0 prices. Anal-

---

[10]If household panel data is available, we can infer price index changes for an individual household.

ogously, we define $P^0(p^0, p^1, y_h^0)$ as the exact price index change between period 1 and period 0 prices holding utility at period 0's level (i.e. $V(p^0, y_h^0) = V(p^1, \frac{y_h^0}{P^0(y_h^0)})$). This price index $P^0(y_h^0)$ converts the household's period 0 nominal income to the hypothetical level of income that would make them equally well off under period 1 prices.[11]

These two price indices are intimately related to equivalent and compensating variation (EV and CV, respectively). $EV_h = e(p^0, u_h^1) - e(p^0, u_h^0) = \frac{y_h^1}{P^1(y_h^1)} - y_h^0$ is the amount of money that would bring a household in period 0 to their period 1 utility, and $CV_h = e(p^1, u_h^1) - e(p^1, u_h^0) = y_h^1 - \frac{y_h^0}{P^0(y_h^0)}$ is the amount of money that would need to be taken away from a period 1 household to bring them back to their period 0 utility.

With this notation in hand, we turn to Lemma 1 (Appendix B contains the formal proof):

Lemma 1. *Assume that prices change over time but relative prices remain unchanged, i.e. $p_i^1 = \lambda p_i^0$ for all i and some $\lambda > 0$.*

   i) *The log price index change for a given income level in period 1, $\log P^1(y_h^1) = \log \lambda$, or a given income level in period 0, $\log P^0(y_h^0) = -\log \lambda$, is equal to the horizontal shift in the Engel curve of any good i at that income level, such that*

   $$E_i(p^1, y_h^1) = E_i(p^0, \tfrac{y_h^1}{P^1(y_h^1)}) \qquad and \qquad E_i(p^0, y_h^0) = E_i(p^1, \tfrac{y_h^0}{P^0(y_h^0)}).$$

   ii) *EV and CV for a given income level are revealed by the horizontal distance along period 1 or period 0's Engel curves, respectively, between the new and old expenditure share, such that $\frac{x_{hi}^0}{y_h^0} = E_i(p^1, y_h^1 - CV_h)$ and $\frac{x_{hi}^1}{y_h^1} = E_i(p^0, y_h^0 + EV_h)$.*

It is easiest to explain Lemma 1 through a simple figure. Figure 2 plots an Engel curve for a particular good in each period. Take as an example a household with initial nominal outlays of y_h^0 (the bottom-left dot in the figure). Since relative prices are not changing, households with the same budget shares must be equally well off as non-homotheticity is the only factor driving relative outlays.[12] Thus, the horizontal distance (in $\log y_h$ space) between their initial position on the period 0 Engel curve and that same budget share on the period 1 Engel curve equals the log of the change in the price index P^0. The CV for this household is then revealed by the additional distance that must be traveled in $\log y_h$ space to go from the crossing point on the period 1 Engel curve to the actual budget share of that household in period 1 (the upper-right dot). The same movements in reverse reveal P^1 and EV.

The proof is simple and relies on homogeneity of degree zero of Marshallian demand and the indirect utility function, i.e. the lack of money illusion. This homogeneity ensures $E_i(p^1, y) = E_i(\lambda p^0, y) = E_i(p^0, y/\lambda)$ when all prices change with a common scalar λ, which here coincides

---

[11]Note that the two price indices are closely related: $y_h^1 = y_h^0/P^0(y_h^0)$ implies $y_h^0 = y_h^1/P^1(y_h^1)$.
[12]Here we abstract from preference (taste) heterogeneity but discuss this possibility in detail in Section 5.2.6.

with the price index change. We can then use Engel curves to infer EV_h from $E_1(p^1, y_h^1) = E_i(p^0, y_h^1/\lambda) = E_i(p^0, y_h^0 + EV_h)$ and CV_h from $E_i(p^0, y_h^0) = E_i(p^1, y_h^0\lambda) = E_i(p^1, y_h^1 - CV_h)$.

Lemma 1 shows that shifts in Engel curves reveal changes in price indices when price changes are uniform across goods. However, if relative prices are unchanged, shifts in Engel curves must be parallel (and price index changes must be identical for households across the income distribution). Motivating Fact 2 clearly shows this is not the case in the Indian context, and is unlikely to be true in other contexts.

To allow for Engel curves consistent with Motivating Fact 2 (i.e. changing slopes over time), we must allow relative prices to change and thus the possibility of income-group specific price indices. However—as we show in Lemma 2 below—if relative prices are allowed to change arbitrarily, Engel curves will not in general reveal changes in price indices and hence welfare.

Lemma 2. *Horizontal shifts in any good i's Engel curve do not recover changes in the log price index under arbitrary changes in the price of good i relative to other goods, or groups of goods.*

Shifts in the Engel curve for good i reflect both changes in utility and changes in relative prices. The proof of Lemma 2 in Appendix B shows that only when demands are Cobb-Douglas will expenditure shares not depend on relative prices, even if in principle we allow such shares to depend on utility. Thus, only when Engel curves are horizontal and fixed (the Cobb-Douglas case) will relative price changes not confound shifts in Engel curves. But in that case, preferences are homothetic and Engel curves are flat, precluding us from identifying price index changes from horizontal shifts. Hence, to be able to relax the assumption of constant relative prices (and thus uniform price index changes across the income distribution), we must impose additional structure on the very general preferences considered in the Lemmas above, and depart from textbook Engel curves.

4.2 Relative Engel Curves and Quasi-Separable Preferences

Lemmas 1 and 2 show that while an appealing concept, shifts in textbook Engel curves will not in general recover changes in price indices when relative prices are changing. To make progress, we propose an approach that departs in two key ways from the previous section—using what we term "relative Engel curves" and assuming that preferences are quasi-separable in the parlance of Gorman (1970, 1976):

Definition *Relative Engel curves, denoted by the function $E_{iG}(p, y_h) = \frac{x_{hi}}{x_{hG}}$, describe how relative expenditure shares within a subset of goods G (i.e. spending on $i \in G$ as a share of total spending on all goods in the group G) vary with log total household outlays per capita.*

Definition *Preferences are quasi-separable in group G of goods if a household's expenditure function can be written as:*

$$e(p, U_h) = \tilde{e}(\tilde{P}_G(p_G, U_h), p_{NG}, U_h) \tag{4}$$

where $\tilde{P}_G(p_G, U_h)$ is a scalar function of utility U_h and the vector of the prices p_G of goods $i \in G$, and is homogeneous of degree 1 in prices p_G.

Quasi-separability is separability in the expenditure function (rather than the utility function). Put another way, quasi-separability imposes no restrictions on substitution patterns between goods within G, or between goods outside of G, but limits substitution patterns between G and non-G goods to operate through a common group-G aggregator. We provide a more detailed primer on quasi-separability below.

   With these two definitions in hand, we turn to the key propositions behind our approach. Proposition 1 makes no assumptions on relative price changes outside of group G—allowing for rich non-homotheticities in the overall price index—but fixes relative prices for goods within G. Proposition 2 then extends these results to allow for relative price changes within G. As we discuss in Section 5.2.2, unlike Proposition 1 which only requires knowledge of expenditures, implementing Proposition 2 requires price data for the subset of goods within G. But crucial in both cases is that we do not require prices for hard-to-measure categories outside of G such as manufactures and services in order to recover the full price index change for all consumption.

Proposition 1. *The following three properties hold for any realization of prices leaving relative prices within group G unchanged (i.e. $p_i^1 = \lambda_G p_i^0$ for all $i \in G$ and for some $\lambda_G > 0$) if, and only if, preferences are quasi-separable in the subset G:*

   i) *The log price index change for a given income level in period 1, $\log P^1(y_h^1)$, or a given income level in period 0, $\log P^0(y_h^0)$, is equal to the horizontal shift in the relative Engel curve of any good $i \in G$ at that income level, such that*

$$E_{iG}(p^1, y_h^1) = E_{iG}(p^0, \tfrac{y_h^1}{P^1(y_h^1)}) \qquad and \qquad E_{iG}(p^0, y_h^0) = E_{iG}(p^1, \tfrac{y_h^0}{P^0(y_h^0)}).$$

   ii) *EV and CV for a given income level are revealed by the horizontal distance along period 0 or period 1's relative Engel curves, respectively, between the new and old expenditure share, such that $\frac{x_{hi}^1}{x_{hG}^1} = E_{iG}(p^0, y_h^0 + EV_h)$ and $\frac{x_{hi}^0}{x_{hG}^0} = E_{iG}(p^1, y_h^1 - CV_h)$.*

   Figure 3 illustrates Proposition 1 graphically, and Appendix B provides the proof. The figure is similar to that for Lemma 1, except now the vertical axis is the relative budget share within G and the two curves are no longer parallel. A similar logic applies, with the horizontal distance from the initial budget share of household h on the period 0 relative Engel curve to the same budget share on the period 1 curve revealing the log of the change in the price index P^0. As before, the additional horizontal distance traveled from that crossing point to the new budget share of that household along the period 1 Engel curve reveals the compensating variation CV_h.

   However, unlike in Lemma 1, since the curves are no longer parallel, the change in the price index P^0 and CV_h may differ depending on the household's position in the income distribution.

14

Relatedly, if we obtain the change in the price index P^1 and EV_h from the shift for the same household but in the other direction, these numbers will no longer be identical to P^0 and CV_h.

Why are the curves no longer parallel? As relative prices within G are held fixed, it is the changes in prices outside of group G (e.g. prices of manufactures and services) that rotate the curves apart if these goods are consumed differentially by rich and poor households. As the proof of Proposition 1 shows, the key role of quasi-separability is that it ensures that these outside-G price changes only affect within-G relative expenditures through changing utility and not through direct price effects. Thus, shifts in relative Engel curves reveal changes in the price index at different points of the income distribution.

To make these statements precise, we highlight several steps of the proof of Proposition 1. Under quasi-separability, relative expenditure in good i within group G can be written as a compensated function $H_{iG}(p_G, U)$ of utility and relative prices within group G only (see Lemma 3 below). This ensures that, holding relative prices within G constant, expenditure shares within group G only depend on utility. The second step is the link from this unobserved compensated Hicksian demand function to observed relative Engel curves by substituting in the indirect utility function $V(p, y)$ that links total outlays and utility:

$$E_{iG}(p^t, y_h^t) = H_{iG}(p_G^t, U_h^t) = H_{iG}\big(p_G^t, V(p^t, y_h^t)\big). \tag{5}$$

In the third step, we show how horizontal shifts in Engel curves in $\log y_h$ space identify changes in price indices. For example, to obtain $P^1(p^0, p^1, y_h^1)$, start with the period 1 relative budget share on the relative Engel curve in period 1:

$$
\begin{aligned}
E_{iG}\big(p^1, y_h^1\big) &= H_{iG}(p_G^1, V(p^1, y^1)) \\
&= H_{iG}(p_G^0, V(p^1, y^1)) \\
&= H_{iG}\big(p_G^0, V(p^0, y_h^1/P^1(p^0, p^1, y_h^1))\big) \\
&= E_{iG}\Big(p^0, y^1/P^1(p^0, p^1, y_h^1)\Big).
\end{aligned}
$$

Equality between the first two lines is an implication of the homogeneous price change $p_i^1 = \lambda_G p_i^0$ within group G (note that H_{iG} is homogeneous of degree zero in prices p_G^t within group G). Equality between the second and third lines follows from the definition of $P^1(p^0, p^1, y_h^1)$ above. The final line simply moves back to relative Engel curves. Thus, the difference between household h's total outlays in period 1 and the total outlays of a household in period 0 with the same relative budget share as h had in period 1 reveals the price index change $P^1(p^0, p^1, y_h^1)$.

Proposition 1 is a strong result. It states that, in theory, we can infer changes in exact price indices and welfare at any given point of the initial or final income distribution by observing: i) relative expenditure shares across some subset G of goods, and ii) total outlays. It also states

15

that this is true if, and only if, preferences are quasi-separable in G, and if relative prices are unchanged within the subset of goods G. By not placing restrictions on relative prices outside of set G, income-group specific price indices can diverge and patterns consistent with Motivating Fact 2 (i.e. non-parallel shifts in Engel curves) can be easily accommodated.

Two further questions naturally arise: how restrictive are the conditions on preferences and on prices in Proposition 1? We address both these issues in turn, first discussing quasi-separability and second relaxing the assumption that prices within G are fixed.

A Primer on Quasi-Separability

To further explore what structure household utility has to possess to satisfy quasi-separability, and discuss which preferences used in the literature fall within this class, we turn to Lemma 3:

Lemma 3. *Preferences are quasi-separable if and only if:*

i) *Relative compensated demand for any good or service i within group G only depends on utility U_h and the relative prices within G:*

$$\frac{x_{hi}}{x_{hG}} = \frac{p_i h_i(p, U_h)}{\sum_{j \in G} p_j h_j(p, U_h)} = H_{iG}(p_G, U_h)$$

for some function $H_{iG}(p_G, U_h)$ of utility and the vector of prices p_G of goods $i \in G$.

ii) *Utility is implicitly defined by:*

$$K\left(\, F_G(q_G, U_h)\,,\; q_{NG}\,,\; U_h\,\right) \;=\; 1$$

where q_G and q_{NG} denote consumption of goods in G and outside G, respectively, for some functions $K\left(F_G, q_{NG}, U_h\right)$ and $F_G(q_G, U_h)$, where $F_G(q_G, U_h)$ is homogeneous of degree 1 in q_G.

This lemma draws on existing results, but Appendix B provides a more direct proof than in previous work. In particular, the equivalence between quasi-separability and condition (i) is shown in Blackorby et al. (1978), and both McFadden (1978) and Deaton and Muellbauer (1980) prove equivalence (ii).

The equivalence with condition (i) is central to our approach and is discussed in the sketch of the proof of Proposition 1 above. In the absence of reliable price data for certain categories of consumption, quasi-separability provides the minimal restriction on preferences such that those unknown prices do not confound shifts in (relative) Engel curves (i.e. that relative consumption only reflects utility and prices within G).

What type of preferences satisfy quasi-separability? With condition (ii), one can see that the

preferences used in Comin et al. (2015) and Matsuyama (2015),[13] in which utility is implicitly defined by $\sum_{i}^{N} \left(\frac{q_i}{g_i(U)} \right)^{\frac{\sigma-1}{\sigma}} = 1$, are quasi-separable in any subset of goods. Using condition i), we can also see that Translog (in expenditure functions), EASI (Lewbel and Pendakur, 2009) and PIGLOG demand systems satisfy quasi-separability in a group of goods G if there are no cross-price effects between goods within and outside of G. Beyond these special cases, condition ii) indicates that we can construct highly flexible demand systems that allow for flexible substitution effects within group G (captured by function F_G) and between goods within G and outside G (function K). For example, we do not need to impose that (Allen-Uzawa) price elasticities are constant across goods within G as in Comin et al. (2015).

The properties of quasi-separable preferences mimic those of direct separability in the dual. However, directly-separable preferences are in general not quasi-separable, and vice versa. Finally, note that quasi-separable demand systems can have any rank in the sense of Lewbel (1991) and so accommodate highly non linear Engel curves consistent with Motivating Fact 1.

Relative Price Changes within G

To relax the assumption that relative prices within G remain unchanged, we need to adjust relative Engel curves to account for the response of within-G expenditure shares to relative prices, holding utility constant. We can then, once again, infer changes in price indices from horizontal shifts in these adjusted curves. Here, we develop this extension formally. Section 5.2.2 lays out an alternative approach if price data are not available for a subset G of consumption.

Proposition 2. *If preferences are quasi-separable in the subset G of goods, the log price index change for a given income level in period 1, $\log P^1(y_h^1)$, is such that*

$$E_{iG}(p^0, \frac{y_h^1}{P^1(y_h^1)}) = E_{iG}(p^1, y_h^1) + \sum_{j \in G} \int_{\log p_j^1}^{\log p_j^0} \frac{\partial H_{iG}}{\partial \log p_j} d \log p_j \qquad (6)$$

where $\frac{\partial H_{iG}}{\partial \log p_j}$ is evaluated along the indifference curve at period 1 utility $U_h^1 = V(p^1, y_h^1)$. Switching superscripts 0 and 1 provides the log price index change $\log P^0(y_h^0)$.

This proposition describes how to adjust relative Engel curves to account for vertical shifts due to changes in within-group-G relative prices.[14] These adjustments require some knowledge of the within-group demand structure H_{iG} and within-group relative price changes. But crucially, they do not require any information on the structure of preferences or prices for goods outside G. As long as there is a group G of goods for which preferences are quasi-separable and reliable price data are available, then relative Engel curves can be adjusted to account for relative price changes and shifts in these curves reveal changes in price indices and welfare.

---

[13]As well as in Fally (2018) where the price elasticity $\sigma(U)$ varies with utility.
[14]Note that Proposition 2 also holds in logs, using $\log E_{iG}$ and $\log H_{iG}$ instead of E_{iG} and H_{iG}.

To be more precise, these vertical adjustments of relative Engel curves depend on compensated changes in expenditure shares within G, holding utility constant. While utility is not directly observed, one can infer compensated changes in within-group expenditures from a Slutsky-type decomposition involving slopes of relative Engel curves and uncompensated price elasticities of within-group expenditure shares (see Lemma 4 in Appendix B):

$$\frac{\partial H_{iG}}{\partial \log p_j} = \frac{\partial E_{iG}}{\partial \log p_j} + E_{jG} \frac{x_G}{y} \frac{\partial E_{iG}}{\partial \log y}.$$

As discussed in detail in Section 5.2.2 below, we implement Proposition 2 in two different ways: as a first order approximation, evaluating each integral as $\frac{\partial H_{iG}}{\partial \log p_j} \Delta \log p_j$, and in its exact form after specifying demand for within-group expenditures H_{iG}.

5 From Theory to Estimation

In this section, we build on the theoretical results above to derive an empirical methodology for estimating exact price indices and changes in household welfare using expenditure survey microdata. We then turn to identification and derive six sets of corollaries to our theoretical propositions that describe testable conditions for unique and unbiased identification. In addition to being of interest in their own right, we draw on these results in our applications in Section 6 to perform a number of validation exercises and robustness checks.

5.1 Estimation Approach

Suppose that we want to estimate the welfare change between two periods for a household with income y_h^0 in reference period 0 and y_h^1 in period 1. The graphical exposition of Proposition 1 in Figure 3 guides a simple estimation approach. First, we use non-parametric methods to estimate flexible relative Engel curves. Importantly, these curves are estimated separately for both periods and for each market, rather than pooling all locations and periods as in existing approaches discussed in Section 2 based on textbook Engel curves. We can then recover changes in exact income-group specific price indices as well as household welfare from the horizontal shift in these curves at different points of the income distribution. Repeating this procedure for multiple goods generates many estimates that can be combined to increase precision (and accommodate unobserved good-specific price and taste shocks as we discuss below).

The first step is to use expenditure survey microdata to estimate kernel-weighted local polynomial regressions of relative expenditure shares, x_{ihm}^t / x_{Ghm}^t, on total outlays per capita for every good i, period t and market m. This provides estimates of $\widehat{x_{ihm}^t / x_{Ghm}^t}$ for households h across the income distribution. Since the Indian NSS data we use in our applications are not a household panel, we use h to index the percentile of the income distribution and explore changes in price indices and welfare at different percentiles. Accordingly, we estimate relative Engel curves

at 101 points corresponding to percentiles 0 to 100 of the local income distribution.[15]

With these relative Engel curves in hand, consider estimating the exact log price index change for income percentile h in period 1, $\log P^1(p^0, p^1, y_h^1)$ (i.e. the price index change holding utility at period 1's level). The relative Engel curve for period 1 provides a point estimate of relative expenditures for households at this percentile of the initial income distribution, x_{ihm}^1/x_{Ghm}^1.

The next step is to estimate the period 0 income level $\widehat{E_{iG}}^{-1}(p^0, x_{ihm}^1/x_{Ghm}^1)$ associated with this relative expenditure share from the crossing point on the period 0 relative Engel curve. To do so, we find the crossing point $\widehat{x_{ih'm}^0/x_{Gh'm}^0}$ and take the corresponding income $\widehat{\log y_{h'}^0}$.[16] As we discuss next, we will restrict attention to monotonic relative Engel curves which ensures this crossing point is unique.

Given these estimates, the income-group specific price index change $\log P^1(p^0, p^1, y_h^1)$ is equal to the difference between $\log y_h^1$ (the period 1 level of income for h) and the estimate of $\widehat{\log y_{h'}^0}$—this is the horizontal shift labeled $\log P^1$ in Figure 3. The welfare change for income-group h, as measured by the equivalent variation, is recovered from the relationship $\log(1 + EV_h/y_h^0) = \widehat{\log y_{h'}^0} - \log y_h^0$, where y_h^0 is the observed period 0 level of income for income percentile h. This expression recovers welfare changes for a hypothetical household that stays at the same point of the income distribution in both periods (if household panel data are available, we could recover welfare changes for a specific household using this methodology).

To estimate the exact price index change holding utility at period 0's level, $\log P^0(p^0, p^1, y_h^0)$, we follow the same procedure but going in the opposite direction (and recovering compensating variation from $\log(1 - CV_h/y_h^1) = \widehat{\log y_{h'}^1} - \log y_h^1$). Each good $i \in G$ provides a separate estimate for $\log P^0$, $\log P^1$, CV_h and EV_h. As discussed in Section 5.2 below, we combine these estimates by taking an average across $i's$ at each percentile of the income distribution.[17]

5.2 Identification

In this subsection, we derive six sets of corollaries (Subsections 5.2.1-5.2.6) related to unique and unbiased identification when taking Propositions 1 and 2 to the data.

5.2.1 Invertibility of Relative Engel Curves

The first result derives necessary and sufficient conditions under which relative Engel curve functions are invertible, and hence our price indices are identified.

Corollary. *Under the same conditions as Proposition 1:*

---

[15]We first smooth the distribution of local income using a local polynomial regression of nominal total outlays per capita on outlays rank divided by the number of observed households. For the relative Engel curves, we use an Epanechnikov kernel with a bandwidth equal to one quarter of the range of the income distribution in a given market. In our applications, we present results using alternative bandwidth choices.

[16]In principle one could estimate Engel curves at many points of the income distribution and find a close-to-exact match. In practice, we take the two closest percentiles and linearly interpolate between them to obtain $\widehat{\log y_{h'}^0}$.

[17]Ultimately, we will use the median as an unbiased estimate of the mean since not all goods $i \in G$ have overlapping relative Engel curves for a particular percentile (see Section 5.2.5).

i) The necessary condition to recover unique estimates of changes in exact price indices and welfare is that different levels of household utility map into unique vectors of relative budget shares within the subset of goods G at any given set of prices.

ii) A sufficient condition for i) to hold is that the relative Engel curve $E_{iG}(p, y_h)$ is monotonic for at least one good $i \in G$.

The first condition is weaker than the second. The practical advantage of the second is that it is readily verifiable in the data, and turns out to be true empirically for all markets and time periods we consider in our applications. For the good-by-good estimation approach that we outline in Section 5.1 above, we restrict attention to good-market combinations where relative Engel curves are monotonic (and so estimates of shifts are unique for each good-market combination). Condition (ii) ensures that we have at least one estimate for each of these markets.[18]

5.2.2 Relative Price Changes Within G

In this section we provide three corollaries that tell us how to use Proposition 2 to account for relative price changes within G groups. First, we show that our estimates are unbiased if an orthogonality condition on the realization of these relative price changes holds. In contexts without any reliable price data, this result provides the identifying assumption necessary for unbiased estimation. In settings—such as ours in Section 6 below—where reliable price data are available for subsets of household consumption, this condition is testable. Second, we show that the orthogonality condition itself serves as a first-order correction term for confounding relative price effects. Third, we show that with additional structure on demands within group G we can also construct an exact correction term.

Orthogonality Conditions and First-Order Correction Terms

Taking a first-order approximation—i.e. assuming that the vertical shifts in relative Engel curves due to within-G relative price changes are proportional to those price changes—and inverting the relative Engel curve at the period 1 expenditure share, equation (6) yields:

$$\log E_{iG}^{-1}\left(p^0, \frac{x_{ih}^1}{x_{Gh}^1}\right) \approx \log\left(\frac{y_h^1}{P^1}\right) + \left(\beta_{ih}^0\right)^{-1} \sum_{j \in G} \sigma_{ijh}(\Delta \log p_j - \overline{\Delta \log p_G}) \tag{7}$$

where $\beta_{ih}^0 = \frac{\partial \log E_{iG}}{\partial \log y_h}$ denotes the slope of the relative Engel curve (i.e. the income elasticity) evaluated at income level y_h^1/P^1 and the initial set of prices p^0, $\sigma_{ijh} = \frac{\partial \log H_{iG}}{\partial \log p_j}$ is the compensated price elasticity of relative consumption of i with respect to price j, and $\overline{\Delta \log p_G}$ is the

---

[18]Specifically, as non-parametrically estimated Engel curves are often noisy at the extreme tails where there are few households across large ranges of outlays, we restrict attention to good-market combinations where Engel curves in both periods are monotonic between percentiles 5 and 95 and drop relative expenditure share estimates beyond those percentiles in cases where those portions are non-monontonic (replacing those values with a linear extrapolation from the monotonic portion of the curve).

average price change within G. The first term on the right-hand side of (7) is the object that we are trying to estimate. The second term is the potential confounder: the vertical shift in relative Engel curves due to relative price changes within G. Even if the bias is large for a specific good, the average bias may be small. Solving for the income that holds utility constant across price environments in (7) and then averaging across $i \in G$, we obtain:[19]

$$\log\left(\frac{y_h^1}{P^1}\right) \approx \frac{1}{G}\sum_{i \in G}\log\widehat{E_{iG}}^{-1}\left(p^0, \frac{x_{ih}^1}{x_{Gh}^1}\right) - \frac{1}{G}\sum_{i,j \in G}\left(\beta_{ih}^0\right)^{-1}\sigma_{ijh}(\Delta\log p_j - \overline{\Delta\log p_G}) \tag{8}$$

We are now in a position to define the following orthogonality condition:

Corollary. *Assuming quasi-separability of subset G, to identify $\log P^1(p^0, p^1, y_h^1)$ in the absence of price data for $i \in G$, relative price changes across i within G must be orthogonal to $\left(\beta_{ih}^0\right)^{-1}\sigma_{ijh}$. Analogously for $\log P^0(p^0, p^1, y_h^0)$ and $\left(\beta_{ih}^1\right)^{-1}\sigma_{ijh}$.*

If relative prices are changing within G but unobserved, we can still recover unbiased estimates of price indices and welfare as long as the within-G price effects are not systematically related to the inverse of the (local h-specific) slopes of the relative Engel curves. Such a condition is informative in contexts where no reliable price data are available for any product group. In these scenarios, price index measurement is extremely challenging. But absent strong priors that relative prices of income elastic goods changed more than inelastic ones within the subset G, our methodology can still provide potentially unbiased estimates. For example, if we are interested in the impacts of shocks or policies that differ by product, such as tariff changes, our price index estimates are likely to be unbiased if the tariff changes are unrelated to relative Engel slopes (a condition that is testable even if reliable price data are not available).

If reliable price data are available for some, but not necessarily all, groups of goods G, this orthogonality condition is testable. In our application, we focus on the 136 food and fuel goods with reliable price data for this reason. If the orthogonality condition is rejected for these G groups, then our price index estimates will be biased. However, the derivation above also provides a simple first-order correction term that corrects for this bias. In particular, we can calculate and add the slope-to-price-change correlation term in equation (8) to our price index estimate. For example, if we further assume a constant elasticity of substitution σ_G within group G, the (market-percentile-level) bias correction term for $\log P^1$ is:

$$-\frac{1}{G}\sum_{i \in G}\left(\beta_{ih}^0\right)^{-1}\sigma_G\left(\Delta\log p_i - \overline{\Delta\log p_G}\right). \tag{9}$$

Thus, as long as we have reliable price data for some G, we can account (to the first order) for confounding within-G price changes and obtain unbiased price index and welfare estimates.

---

[19]Symmetrically for P^0: $\log\left(\frac{y_h^0}{P^0}\right) \approx \frac{1}{G}\sum_{i \in G}\log\widehat{E_{iG}}^{-1}\left(p^1, \frac{x_{ih}^0}{x_{Gh}^0}\right) - \frac{1}{G}\sum_{i,j \in G}\left(\beta_{ih}^1\right)^{-1}\sigma_{ijh}(\Delta\log p_j - \overline{\Delta\log p_G})$.

Exact Correction Terms

An exact correction may be preferred to the first-order approximation above if within-G relative price changes are large. Proposition 2 tells us that knowing the shape of the function $H_{iG}(p_G, U)$ is sufficient to compute an exact adjustment to account for within-group relative price changes. A case that yields a particularly simple (but exact) adjustment term is the iso-elastic case, assuming a within-group structure akin to Comin et al. (2015): $H_{iG}(p_G, U) = \frac{A_i(U)p_i^{1-\sigma}}{\sum_{j \in G} A_j(U)p_j^{1-\sigma}}$. If we have estimates of the single price elasticity σ, it is then straightforward to calculate the price index change as the horizontal shift between the actual relative Engel curve in one period and the counterfactual (i.e. relative-price adjusted) relative Engel curve in the other period. For example, to calculate P^1, we would calculate the counterfactual period 1 relative Engel curve, $\widetilde{x_{ih}^1/x_{Gh}^1}$, as follows:

$$\widetilde{x_{ih}^1/x_{Gh}^1} = \frac{(p_i^1/p_i^0)^{1-\sigma} \, (x_{ih}^1/x_{Gh}^1)}{\sum_{j \in G}(p_j^1/p_j^0)^{1-\sigma} \, (x_{jh}^1/x_{Gh}^1)}, \tag{10}$$

before calculating the horizontal distance between this curve and the actual period 0 relative Engel curve, i.e. equating $E_{iG}(p^0, \frac{y_h^1}{P^1(y_h^1)}) = \widetilde{x_{ih}^1/x_{Gh}^1}$.

Alternatively, a simple additive specification has constant semi-elasticities ξ within group G akin to EASI demands (Lewbel and Pendakur, 2009). In this case, the adjustment is as follows:

$$\widetilde{x_{ih}^1/x_{Gh}^0} = \frac{x_{ih}^1}{x_{Gh}^1} + \xi \times (\Delta \log p_i - \overline{\Delta \log p_G}). \tag{11}$$

5.2.3 Aggregation Across Varieties of a Good

Typically, researchers estimate Engel curves for a "good" (indexed here by g) that itself contains potentially many sub-varieties (is in the exposition above, e.g. different preparations, brands, pack-sizes or flavors), either because that is the level the data are reported at or because specific varieties are not consumed widely enough given the number of households sampled. Fortunately, Propositions 1 and 2 can also apply to aggregates of varieties of a good rather than individual varieties, even if demands for those varieties are non-homothetic within g. For instance, imagine there are fancy packaged sea salts and simple table salt that are consumed in different proportions by the rich and the poor and sold in different types of shops.

Corollary. *Suppose that G in our exposition above can be partitioned into subgroups of goods: $G = g1 \cup g2 \cup g3...$ (e.g. salt, meat, vegetables etc.). Denote by $E_{g,G}$ the expenditure share on subgroup g within group G. Under the assumptions of Proposition 1:*

$$E_{g,G}(p^1, y_h^1) = E_{g,G}(p^0, \frac{y_h^1}{P^1(y_h^1)}) \qquad and \qquad E_{g,G}(p^0, y_h^0) = E_{g,G}(p^{t1}, \frac{y_h^0}{P^0(y_h^0)}).$$

In other words, the key equivalence in Proposition 1 continues to hold if we treat the sub-

groups g as products (instead of the individual varieties i). Furthermore, under the assumption that prices within each subgroup g can be aggregated across the i into price indices, $P_g(p_g, U)$, we can apply Proposition 2 and the price-adjustment corollaries above to correct for relative price changes, but now using subgroup price indices $P_g(p_g, U)$ instead of individual prices p_i.[20]

Several remarks are in order. First, note that these subgroup price indices can be non-homothetic: relative consumption within subgroup g can vary with utility U (and thus income); the rich and poor can even consume distinct varieties. Second, aggregation can accommodate differences in shopping amenities and store-level price differences (modeled as store-specific varieties). Third, aggregation can accommodate new and disappearing varieties within subgroup g using existing methods (e.g. Feenstra, 1994). Finally, a more practical advantage is that relative Engel curves for subgroup g may be strictly monotonic while consumption of specific varieties is zero (and thus flat) for some periods and ranges of income.

Taken together, these aggregation results are particularly valuable when implementing our approach to estimate price indices and welfare from highly-disaggregated real-world data—e.g. barcode-level consumption data—that are available for some subset of consumption G.

5.2.4 Quasi-Separability and Misclassification

Using Lemma 3 above, we provide a test for quasi-separability using expenditure survey data.

Corollary. *If, and only if, preferences are quasi-separable in group G, the price elasticity of the uncompensated (i.e. holding income fixed) expenditure share $x_{iG} \equiv \frac{x_i}{x_G}$ in the price of any good $j \notin G$ equals the slope of the relative Engel curve multiplied by good j's overall budget share:*

$$\left. \frac{\partial \log x_{iG}}{\partial \log p_j} \right|_y = -\frac{p_j q_j}{y} \frac{\partial \log x_{iG}}{\partial \log y}.$$

In the formulation above, $\left. \frac{\partial \log x_{iG}}{\partial \log p_j} \right|_y$ corresponds to the vertical shift of the relative Engel curve induced by the change in the price of good j. This can be tested if reliable price data are available for some goods outside subset G. Alternatively, we can use such data to explore the horizontal shift induced by this price change which is equal to the ratio $\left. \frac{\partial \log x_{iG}}{\partial \log p_j} \right|_y / \frac{\partial \log x_{iG}}{\partial \log y}$. Under quasi-separability, this ratio coincides with the marginal effect of a good j price-change on the price indices P^0 and P^1. This result generates a second and more straight-forward test:

Corollary. *The elasticity of the exact price index P^t, $t \in \{0, 1\}$ with respect to the price of any good j equals the overall expenditure share of good j:*

$$\frac{\partial \log P^t}{\partial \log p_j^t} = \frac{p_j^t q_{jh}^t}{y_h^t}. \tag{12}$$

---

[20]For example, the price aggregates derived in Redding and Weinstein (2020) could be used for $P_g(p_g, U)$, assuming within-g preferences have their CES structure.

This equality is simply Shephard's Lemma applied to our price indices. Since we do not use prices from non-G goods to estimate our price indices, our estimation strategy does not guarantee that this equality holds. One additional benefit of this test is that it also serves as a smell test for our general approach. Recall that we are calculating a price index covering the full consumption bundle from relative expenditures within some group G. The test asks whether our estimated price index responds to price changes for goods outside group G as it must (and will only do so fully if quasi-separability holds).

Misclassification Bias

We provide two further results related to quasi-separability that guide our choice of the subset G in the empirical application. First, what if we misclassify a good i that truly belongs in G as a non-G good (i.e. we omit a good that belongs within G)? Now a relative price shock for this omitted good may shift relative Engel curves within G, even if we hold utility constant. Approximating to the first-order, the bias due to misclassifying product 0 as non-G is:

$$\frac{1}{G}\sum_{i\in G}\log E_{iG}^{-1}\left(p^0,\frac{x_{ih}^1}{x_{Gh}^1}\right) \approx \log\left(\frac{y_h^1}{P^1}\right) + (\Delta\log p_0 - \overline{\Delta\log p_G}) \times \frac{1}{G}\sum_{i\in G}\left(\beta_{ih}^0\right)^{-1}\sigma_{i0h} \quad (13)$$

(where σ_{ijh} denote semi-elasticities). We can see that there is no additional bias from omitting good 0 if either of the following conditions hold: i) the price change of good 0 equals the average change in prices among G goods, and ii) if the price elasticity with respect to good 0 is small or uncorrelated with Engel curve slopes, i.e. $\frac{1}{G}\sum_{i\in G}\left(\beta_{ih}^0\right)^{-1}\sigma_{i0h} = 0$.

Second, what if we misclassify a non-G good j as part of G? To the first order, the bias is:

$$\log E_{jG}^{-1}\left(p^0,\frac{x_{jh}^1}{x_{Gh}^1}\right) \approx \log\left(\frac{y_h^1}{P^1}\right) + \left(\beta_{jh}^0\right)^{-1}\sum_{k\in G,NG}(\Delta\log p_k - \overline{\Delta\log p_G})\frac{\partial\log H_{jG}}{\partial\log p_k}, \quad (14)$$

where $H_{jG}(p,U)$ is the expenditure share on j within G, which now depends on all prices. Hence, there is no bias if the relative price change across all goods k is not correlated with the difference in (compensated) price elasticities between j and G (or, of course, if there is no change in relative prices as in Lemma 1). These two results motivate assessing the potential biases from violations of quasi-separability due to misclassification by exploring sensitivity to the categorization of goods into G groups.

5.2.5 Unobserved Welfare Changes (Sample Selection)

Not all levels of household utility in period 0 are necessarily observed in period 1 and vice versa. For example, when evaluating price index changes P^0 for poor households in period 0, there may be no equally poor households in period 1 if there is real income growth (and similarly when evaluating P^1 for rich households in period 1). This means that Engel curves may not

always overlap in budget share space (i.e. horizontally) for all income percentiles, and this gives rise to sample selection concerns, especially at the tails.

These selection issues take two forms, missing goods and missing markets. Recall from Section 5.2.2 that averaging multiple price index estimates (one for each good for which we can calculate the horizontal shift in relative Engel curves) at a particular percentile h can potentially eliminate bias from relative price shocks within the G group (or taste shocks as we discuss below). However, in the presence of such shocks (vertical shifts), some goods may have overlap and others may not. Averaging over only the subset of goods for which there is overlap at a given percentile h generates potential biases since overlapping and non-overlapping goods experienced different shocks. This is particularly problematic at the tails of the distribution. For example, if there is no true overlap when estimating P^0 for poor households, any overlapping goods we do observe must have experienced large vertical shocks to relative Engel curves which would generate large enough price index estimates to make poor households in period 1 appear as poor as they were in period 0.[21]

To address such sample selection concerns, we exploit the fact that we observe whether or not a given good has missing overlap at a given income percentile and whether this good is censored from above or from below (which depends on the sign of the slope of the relative Engel curve). Combining this information with the identifying assumption that the distribution of price index estimates across different goods within G is symmetric for a given income percentile allows us to consistently estimate the price index change. To do so, we order the observed (i.e. overlapping goods) and unobserved (i.e. non-overlapping goods) price index estimates and take the median (which is an unbiased estimate of the mean).[22] In cases where the median is unobserved, we can impose a stronger assumption: that the distribution of price index estimates across different goods within G is uniform for a given income percentile. That allows us to solve for the mean as long as at least two goods overlap (see Sarhan, 1955). In our Indian application below, symmetry alone appears sufficient to solve selection issues.

A different type of sample selection arises if we don't observe any relative Engel curves that overlap for a given percentile and market. In this case, we face a market-level sample selection issue when aggregating across markets. In particular, there may be missing markets among poor percentiles for P^0 and missing markets for rich percentiles for P^1 if real incomes grew. In practice, we find that almost no markets are missing after we implement the good-level selection correction above (i.e. we observe overlap in monotonic relative Engel curves for at least two goods for close to every decile-market pair in our sample). Therefore, the good-level selec-

---

[21]In this example, averaging estimates for overlapping goods would lead to an upward-bias in P^0 relative to the true inflation faced by poor households.

[22]We rank estimates, placing unobserved estimates below the lowest or above the highest estimate depending on whether they were censored from below (e.g. when calculating P^0 for poor households or P^1 for rich households) or above (e.g. when calculating P^0 for rich households and P^1 for poor households).

tion correction is sufficient in our context to solve market-level selection issues. Were it not, we could apply two-step Heckman selection corrections or make assumptions on the distribution of estimates across markets to recover the missing markets for a given percentile h.

5.2.6 Taste Heterogeneity and Taste Changes

Finally, we formally consider three concerns related to taste heterogeneity and taste changes.

Omitted Variable Bias in Engel Curve Estimation

The first issue is common to all Engel curve estimates: if household taste differences are correlated with household income, Engel curves will be biased. For example, more educated households may both value certain goods more and have higher incomes. This would bias our price index and welfare estimates, as changes in real income over time would not affect budget shares in the way our estimated Engel curves predict. This bias can be addressed either by controlling for household characteristics in the estimation of relative Engel curves or by estimating curves separately for different types of household (both of which we pursue in Section 6).

Heterogeneous Price Index Changes

The second issue is that if tastes for goods differ across household types within a given income percentile in a way that correlates with relative price changes over time, then price index and welfare changes for a given income percentile will differ by type. In this case, we show that our method yields a weighted average change: $\frac{y_h^1}{\widetilde{P}^1(y_h^1)}$, where $\widetilde{P}^1(y_h^1) \approx \sum_k w_k^1(y_h^1) P_k^1(y_h^1)$ with weights given by the relative Engel slopes of type k: $w_k \equiv \sum_i (\beta_{i,k}^1/\overline{\beta_i^1}) / \sum_{k'} \sum_i (\beta_{i,k'}^1/\overline{\beta_i^1})$. In this scenario, if one is interested in the welfare change for a particular household type, such as households with large family sizes, we can carry out our procedure just for those households.

Changes in Tastes Over Time

The third issue arises when household tastes change over time. Such taste changes are only problematic if they are systematically related to differences in slopes of relative Engel curves across goods. To be more precise, we can derive an orthogonality condition that is analogous to the orthogonality condition we derive in (8) for unobserved relative price changes within G. Denoting taste shocks—i.e. shifts in budget shares conditional on prices and income—by $\Delta \log \alpha_{ih}$ and abstracting from relative price changes over time, we obtain:

$$\log \left(\frac{y_h^1}{P^1} \right) \approx \frac{1}{G} \sum_{i \in G} \log \widehat{E_{iG}}^{-1} \left(p^0, \frac{x_{ih}^1}{x_{Gh}^1} \right) - \frac{1}{G} \sum_{i \in G} \left(\beta_{ih}^0 \right)^{-1} \Delta \log \alpha_{ih} \ .$$

If taste shocks across i within subset G are orthogonal to the local slope of the relative Engel curve in period 0 (or 1 to identify P^0), the bias averages to zero across goods.[23]

---

[23]To ensure that shares add-up to unity within G, we assume that these taste shocks sum up to zero. These shocks can be defined using shifters multiplying P_G in equation (4) or as price shifters as in Redding and Weinstein (2020).

Unfortunately, such a condition is not in general testable. One scenario that may violate this condition is if household types have different tastes and there are compositional changes over time (e.g. increases in education). This concern will only be problematic if different household types have different price index changes over time, a condition that we can (and do) explicitly test by separately estimating and comparing price index changes for different household types.

6 Applications

In the final section we apply our "New Engel" methodology introduced in the previous sections. We first explore changes in rural Indian welfare over time, and then re-visit the welfare impacts of India's 1991 trade reforms.

6.1 Data Sample, Product Aggregation and Product Groups

We use the Indian NSS microdata described in Section 3 to estimate changes in household price indices and welfare for rural Indians between 1987-87 and 1999-2000. We do this for 9 income deciles (10th–90th percentile) in each district. Given the need to non-parametrically estimate relative Engel curves, we restrict attention to the 249 districts where we observe at least 100 households in both survey rounds. (As we show, results are not sensitive to this restriction.)

The estimation requires multiple products i within quasi-separable groups G. As discussed above, to correct for within-G price changes and test for quasi-separability, we restrict attention to the food and fuel product groups for which we have well-measured prices (as identified by Deaton, 2003b). To reduce measurement error when estimating relative Engel curves for rarely consumed items, we aggregate these 136 products to the second-lowest level of aggregation in the NSS surveys, which yields 34 products indexed by g (see Appendix Table A.1 for further details). The results in Section 5.2.3 prove that such an aggregation is admissible, and that we can implement price corrections, as long as we can measure price indices $P_g(p_g, U)$ for these 34 g goods (see 6.2.1 below). As shown in Appendix Figure A.1, this aggregation dramatically reduces the share of empty market-by-product cells (from 50 percent to less than 15 percent), and moving to the next highest level of aggregation (8 goods) provides little additional benefit.[24]

We divide these 34 aggregate products into three broader consumption groups: raw food products (e.g. rice, leafy vegetables), other food products (e.g. milk, edible oils) and fuels (e.g. firewood, kerosene). In our baseline estimation, we assume these three groups each form a quasi-separable G group, with all remaining goods and services (e.g. processed food, manufactures and services) excluded as part of the NG group. We combine estimates for all three groups by taking medians as described in Section 5.2.5.[25] As we describe below, Figure 8 explores ro-

---

[24]Appendix Figure A.2 reports qualitatively similar inflation estimates using these alternate levels of aggregation.

[25]In principle, comparing estimates obtained from different G groups provides an over-identification test (i.e. price index estimates from different G groups should be identical if quasi-separability and orthogonality conditions on tastes and prices are satisfied). However, given the limited number of products with both monotonic and over-

bustness across 108 perturbations of sensible G groupings.

6.2 Changes in Indian Price Indices and Welfare Over Time

Before describing the results of our approach, we first summarize income and price changes between 1987 and 2000 using the best existing Indian CPI statistics. Figure 4 plots growth rates in nominal household outlays per capita for each decile of the local income distribution (using population-weighted averages of log changes across all 249 rural districts).[26] Growth exceeded 200 percent and there is a clear and strong pattern of nominal income convergence over this 13-year period, with outlays per capita rising substantially faster for the poor than for the rich.

The left panel of Figure 5 plots averages of Paasche and Laspeyres price index estimates using the methodology of Deaton 2003b that draws on observed price changes weighted by district-level expenditure shares for the 136 food and fuels items where price data are deemed reliable.[27] Mechanically, these price indices do not vary across the income distribution, and so the estimated 160-170 percent rise has little bearing on the convergence noted above. The middle panel of Figure 5 relaxes this homotheticity by using district-decile specific expenditure shares when calculating Paasche and Laspeyres price indices. While inflation (as measured by standard food and fuel price indices) was broadly similar across groups, if anything these measures suggest higher inflation for the rich, implying further convergence in real incomes beyond what is apparent from Figure 4.

The right panel of Figure 5 presents our baseline price index estimates using the New Engel approach described in 5.1 (as above, plotting population-weighted averages across districts by decile). We obtain bootstrapped confidence intervals (here, and analogously for nominal outlays and the Paasche and Laspeyres indices above) by sampling with replacement 1000 times from the distribution of households within each district-survey round and plotting the 2.5 and 97.5 percent envelope of price index estimates at each decile.

Two main findings emerge. First, the New Engel approach generates broadly similar estimates of Indian consumer price inflation among low-income deciles compared to existing CPI estimates that are based on changes in observed prices for food and fuel. Since these product groups represent a sizable fraction of rural household consumption for poor households in India (82 percent at the 5th percentile when averaging across both survey rounds), this finding is reassuring—particularly since we are comparing a standard price index that explicitly uses observed price changes to one that only uses changes in expenditure shares in conjunction with

---

lapping relative Engel curves within each G in our setting, these orthogonality conditions are unlikely to be satisfied without pooling the estimates.

[26]We report percentage changes for all outcomes (incomes, price indices and welfare) calculated by exponentiating the mean log change across survey rounds, weighting districts by population weights.

[27]Price changes are computed from changes in district median unit values for each of the 136 food and fuel items, with products weighted by mean district-level expenditure shares (where shares are calculated using household survey weights). We replace any missing district-level price changes with state-level ones.

our relative Engel curve estimates (in the next section we do use food and fuel price data, but only to correct for within-G relative price changes).

Second, we estimate that cost of living inflation has been substantially higher for poor Indian households compared to the rich, the opposite of what one would infer from the food and fuel Paasche and Lespeyres indices. Figure 6 combines the estimated changes in nominal incomes and price indices to obtain welfare changes (EV and CV in the New Engel approach, and real income for the standard CPI approach). These differential rates of inflation eliminate any convergence in wellbeing between the rich and poor over this period. In fact, if anything, welfare grew more for rich households.

Why are our New Engel price index estimates lower for richer households? The most likely explanation is that high-income households disproportionately benefited from price falls, new varieties, and quality increases in consumption categories where price measurement is challenging (and so are omitted in Deaton's CPI approach). In particular, the rich spent a large and increasing share of their budget on durables such as manufactures and on services, categories for which unobserved quality differences make price data unreliable.[28] As touched upon in the introduction, lower inflation in these specific categories is consistent with the fact that the Indian trade reforms were centered on manufacturing intermediates which substantially raised the quality and variety of Indian manufactures (Goldberg et al., 2010); and that there was a dramatic increase in share of services in GDP over the reform period (Mukherjee, 2015).

Beyond accounting for inflation in hard to measure categories, our methodology is also immune to concerns that lie at the center of the Great Indian Poverty Debate. India's 1999-2000 NSS added an additional 7-day recall period for expenditures on food products which inflated answers to the consistently asked 30-day recall questions. The most influential solution, that of Deaton (2003a), adjusts food expenditure using the mapping between food expenditure and fuels (for which no additional recall period was added) from earlier rounds. That solution requires that relative prices of food and fuels did not change. In contrast, our estimates are robust to the additional recall period as long as it did not change *relative* consumption shares within a given food or fuel group G. This condition is testable using the thin NSS round 54 (1998) where, in order to test proposed changes to the surveys, households were randomly assigned to to different recall periods. Consistent with our claim, Appendix Table A.2 shows that the choice of recall period did not affect relative consumption shares within our G groups.[29] Thus, our finding of no real income convergence between rich and poor has the potential to inform, and revise the conclusions of, the Great Indian Poverty Debate summarized in Deaton and Kozel (2005).[30]

---

[28]Averaging across rounds, the richest 5th percentile spent 32 percent of their expenditure on these categories.

[29]In addition, Appendix Figure A.3 shows similar patterns of pro-rich inflation between the 1987/88 and 1994/95 survey rounds when the questionnaire was unchanged.

[30]Two additional notes. First, even if relative outlays across products within G were affected by the additional recall period, such effects would still need to be systematically related to slopes of relative Engel curves to bias our

6.2.1 Price Corrections and Validation Results

In this subsection, we perform a number of validation exercises that follow from our corollaries in Section 5.2, as well as reporting several additional robustness checks.

Relative Price Changes Within G

Figure 7 applies the first-order and exact corrections for potentially confounding relative price changes within G groups that we describe in Section 5.2.2. As we focus on food and fuel product groups for which we observe reliable price data, these corrections are straightforward to compute.[31] For the first-order correction term, we assume a common elasticity of substitution of $\sigma = 0.7$ based on averages from Cornelsen et al.'s (2015) systematic review of food price elasticities in low income countries estimated using similar levels of aggregation to our 34 goods. For the exact price correction, we use the isoelastic correction (non-homothetic CES) in equation (10) with the same $\sigma = 0.7$ elasticity assumption. Reassuringly, Figure 7 shows that the estimated inflation rates across deciles change very little after adjusting for relative price changes within G groups. Put another way, recall that our first-order correction term also serves as a test of our orthogonality condition. Thus, the fact the estimates change little implies that relative price changes within our three food and fuel G groups are either small or only weakly related to income elasticities.

Quasi-Separability and Misclassification

We first present the simpler of the two tests of quasi-separability discussed in Section 5.2.4. The test predicts that the elasticity of the price index, calculated using only the subset of goods in G, with prices of an outside good j should equal the expenditure share of the outside good (equation 12). Given that we have reliable price data for only foods and fuels, we implement a test of quasi-separability by re-estimating the log of the price index from food expenditures only (i.e. using only 2 of the 3 Gs) by district and by decile, and regressing these indices on log fuel price changes interacted with fuel expenditure shares.[32] Assuming fuel price changes across districts are independent of other unobserved price changes, we expect a coefficient equal to unity. Note that this test goes to heart of our methodology that recovers the complete price index for all goods despite only using relative consumption for a subset of goods for which we have reliable price data. In particular, it asks whether relative consumption within a particular group (food in this case) successfully captures price changes outside that group (fuels in this case, but more generally manufactures and services where prices are poorly measured).

---

welfare estimates (see Section 5.2.2). Second, while recall bias does not affect our estimates of welfare (EV and CV), the decomposition between nominal income changes and price-index changes is potentially affected.

[31]We use a Fisher price index to aggregate the observed price changes of the 136 products i in the NSS to 34 goods g (using expenditure shares within each aggregate product to compute weights).

[32]Of course, this is only a partial test, as we do not know if quasi-separability holds with respect to goods for which we do not observe prices.

We show the results of this test in Table 2. Panel A uses our baseline New Engel price index estimate calculated using only food groups (i.e. excluding fuels), while Panel B additionally applies the exact correction for relative price changes described above. The first two columns report results for P^0, and the last two for P^1. Columns 2 and 4 include district fixed effects so that we exploit within-market variation across deciles (i.e. do our estimated price indices increase with fuel prices relatively more for deciles with larger expenditure shares on fuel?). In support of our quasi-separability assumption, coefficients are generally close to unity (and in no case can we reject a coefficient of one).

Next, we investigate bias from potential violations of quasi-separability due to misclassifying products into three broad G groups. To this end, we re-estimate our price indices for each decile and market across 108 sensible splits of our g goods into plausibly quasi-separable groupings G.[33] Figure 8 presents the estimation results for each decile, plotting our baseline point estimate on top of the mean and 95th percentile confidence intervals of point estimates from the 108 possible G groupings. Reassuringly, our baseline specification is close to the mean of the 108 estimates for every decile of the income distribution. In addition, the confidence intervals are reasonably tight—suggesting that the conditions under which misclassification bias is small (equations 13 and 14) are met in our setting.

Sample Selection Issues

As described in Section 5.2.5, our baseline estimates address sample selection issues due to non-overlapping relative Engel curves by ranking both missing and non-missing estimates and taking the median under the assumption of a uniform distribution of estimates across $g \in G$. Appendix Figures A.4-A.6 illustrate and assess these sample selection issues. The left panel of Appendix Figure A.4 presents the price index estimates that do not correct for non-overlap issues and simply average over non-missing goods. Compared to our baseline (the right panel), the biggest discrepancies occur for P^0 among the poorest deciles and P^1 among the richest deciles. It is exactly these households for which overlap issues are most severe since, given economic growth, the welfare levels of the poorest households in period 0 are typically not observed in period 1 and similarly for the welfare levels of the richest households in period 1.[34]

The middle panel of Appendix Figure A.4 implements only the first step of our selection correction. Recall that we only require symmetry (not uniformity) of the distribution of price index estimates in order to take medians of the ordered non-missing and missing price index

---

[33]As shown in Appendix Table A.1, the 34 g products fall into three high-level groups (raw food, other food and fuel) and 8 subgroups within those. To discipline plausibly quasi-separable nests G, we impose that a g can only be bundled together with other gs in the same high-level group. Additionally, different gs within one of the 8 subgroups cannot be grouped into more than one G. With these restrictions, we generate 105 possible ways of allocating gs into G groups (i.e. $(2^4 - 1) \times (2^3 - 1) \times 1 = 105$, including our baseline). Finally, we add three more cases: only 1 G group across all products, 2 G groups (food and fuel), and 8 G groups (one for each subgroup above).

[34]Figure A.5 illustrates this fact by showing the frequency of non-overlapping estimates by decile, broken out by type of non-overlap (censored from above or from below) that we use to rank missing estimates.

estimates and obtain unbiased estimates of the mean. Only applying symmetry (middle panel) eliminates almost all the discrepancy between P^0 and P^1 due to sample selection issues and generates very similar estimates to our uniformity baseline (right panel). However, by only imposing symmetry, we lose any market-decile pairs for which the median ranked good has no overlap. Appendix Figure A.6 plots the number of market-decile pairs for which we obtain price index estimates under both symmetry and uniformity. As the distribution of log nominal outlays per capita is right-skewed, Engel curves are less likely to overlap when calculating P^0 (and compensating variation), and so a substantial number of pairs are missing when only imposing symmetry. However, we obtain estimates for essentially all market-deciles once uniformity is imposed and so market-level selection issues do not arise under our baseline specification.

Taste Heterogeneity and Taste Changes

We now investigate concerns that our estimates may be affected by taste heterogeneity across households or taste changes across time (see Section 5.2.6). Appendix Figure A.7 recalculates price indices using non-parametric Engel curves that condition on a standard set of linear controls for household characteristics.[35] Reassuringly, results change little, suggesting that systematic bias in estimates of cross-sectional Engel curves is unlikely to be driving our findings.

Appendix Figure A.8 corroborates this finding by presenting separate price index estimates for different types of rural households; small versus large households, high versus low education, young versus old, and literate versus illiterate (with the last three comparisons based on characteristics of the household head). Recall from Section 5.2.6 that these exercises are informative on a number of fronts. First, by estimating Engel curves separately across demographic groups, we limit potential bias in estimates of cross-sectional Engel curves. Second, we can explore to what extent different types of household experienced different inflation rates conditional on their position in the income distribution (i.e. due to taste heterogeneity). Third, we can address concerns that the composition of household types may have changed over time, biasing estimates if taste heterogeneity across types is systematically related to slopes of relative Engel curves (e.g. if average education or household size changed over time and educated or large households have different tastes). The fact that the price index estimates show very similar patterns for different household types provides reassurance that taste heterogeneity and taste changes (at least those due to compositional changes) are not driving our findings.

Additional Robustness Checks

We report several additional robustness checks. Appendix Figure A.9 presents results for alternative bandwidth choices when non-parametrically estimating relative Engel curves and al-

---

[35]In particular, for each good and market (pooling across both periods) we estimate coefficients on the following controls: a scheduled caste dummy, a literacy of household head dummy, log of household size, and the share of children in the household. We then use relative Engel curves for each good-period-market evaluated at the control's market-level median (i.e. evaluated at a value that is constant over time).

ternative strategies to deal with noise at the tails. Appendix Figure A.10 reports results without restricting attention to markets with at least 100 household observations in both survey rounds. Reassuringly, results are not qualitatively different to our baseline estimates.

6.3 Revisiting the Impacts of India's 1991 Trade Reforms

In this section, we revisit the impact of India's 1991 trade reforms on the welfare of rural households in India. We closely follow Topalova's (2010) analysis that pioneered the (now widespread) use of a shift-share instrument to identify the impacts of trade shocks. We focus on her most stringent specification that regresses poverty head count ratios (the dependent variable, using the Deaton, 2003a recall bias correction discussed above) on district-level exposure to import tariff cuts (the independent variable) for rural districts across the 1988/89 and 1999/2000 NSS rounds. Exposure is measured as the weighted average tariff cut, with weights proportional to the initial-period sectoral employment shares in the district. She also includes district fixed effects, time fixed effects, and several additional time-varying district controls.[36] Two variables are used to instrument the potentially-endogenous shift-share tariff exposure measure. The first is the same shift share measure but calculated only using tradable industries (to remove variation due to differences in industrial employment shares across districts). The second instrument uses this tradable shift share but now using the initial average level of import tariffs rather than the change (as all tariffs were brought to similar levels post reform, initially higher tariffs fell more for predetermined reasons).

We revisit this regression but replace the outcome (district-level rural poverty rates) with our estimated welfare metrics.[37] Importantly, our method allows us to calculate impacts at each decile of the local income distribution. The right panel of Figure 9 plots the decile-specific coefficients on the tariff exposure variable (i.e. the difference in welfare growth for more exposed regions compared to less exposed). For expositional purposes, we focus on our equivalent variation welfare metric.[38] For comparison, the left panel plots estimates for the same specification but replacing the dependent variable welfare with log nominal outlays per capita.

Two main findings emerge. First, while existing work has focused on the effect on poverty rates, our estimation reveals that the adverse effects of import competition on local labor markets are borne by households across the income distribution, including by rural households in the richest income deciles. Second, we find that the adverse effects on nominal outcomes are amplified when taking into account the effects on household price indices. Import competition leads to relatively higher local price inflation, particularly for richer households. This somewhat

---

[36]This specification corresponds to column 8 in Table 3a) of Topalova (2010).

[37]We obtain similar results using Topalova's other specifications although, as in Topalaova, results are less significant. Note that Topalova does not restrict attention to markets with more than 100 survey households. Restricting Topalova's sample in this way makes her effect sizes larger.

[38]As discussed above, we have more overlapping Engel curves and so less noise when calculating P^1 and EV.

surprising finding is not simply an artifact of our approach, as it also appears when calculating a simple Laspeyres index using the raw price data for food and fuels (see Appendix Figure A.11). One potential explanation is that hard hit areas did not experience the same increases in retail-sector competition or productivity as faster-growing areas. An alternative explanation, and one beyond the scope of this paper, is that the shift-share exclusion restriction is violated.

7 Conclusion

Measuring changes in household welfare and the distribution of those changes is challenging and typically requires the researcher to observe the full vector of quality- and variety-adjusted price changes—an incredibly difficult task for categories such as manufacturing and services. In this paper, we propose and implement a new approach to estimate changes in household price indices and welfare across the income distribution from horizontal shifts in relative Engel curves. In poor data environments without any reliable price information, we prove that if relative price changes within some quasi-separable product group are unchanged (or uncorrelated to the slope of relative Engel curves) but price changes outside this group are unrestricted, such an approach uncovers theory-consistent and exact price indices as well as welfare changes, despite only drawing on widely-available expenditure survey data. Where reliable price data do exist for some subset of goods, we can relax the restrictions on relative price changes within the quasi-separable group, as well as validate our assumptions on preferences. Compared to existing Engel approaches (e.g. Hamilton, 2001), we allow for non-homothetic price index changes, relax the implicit need to observe the full vector of price changes, and also accommodate arbitrarily non-linear Engel curves.

We apply this new method to measure changes in household welfare and revisit the effects of trade over India's reform period. We find that consumer price inflation was substantially higher for poor households than rich, essentially eliminating the convergence seen in nominal incomes. This finding is missed by standard price indices using the subset of consumption where prices are well measured (and is consistent with falling prices and rising quality/variety in manufactures and services which are both disproportionately consumed by the rich and hard to measure prices for). Second, going beyond poverty rates, our estimation reveals that the adverse effects of import competition on local labor markets in India are borne by households across the entire income distribution, including the richest.

Beyond providing a deeper understanding of India's economic reforms, we believe our methodology is widely applicable in the many settings where expenditure survey data are available or can be easily collected. Given the increasing availability of survey microdata over time and across space, and the growing interest in distributional analysis, the usefulness of such an approach is only likely to grow.

References

ALMÅS, I. (2012): "International income inequality: Measuring PPP bias by estimating Engel curves for food," *American Economic Review*, 102(2), 1093–1117.

ALMÅS, I., T. BEATTY, AND T. CROSSLEY (2018): "Lost in translation: What do Engel curves tell us about the cost of living?," *IFS working paper W18/04*.

ALMÅS, I., AND A. KJELSRUD (2017): "Rags and riches: Relative prices, non-homothetic preferences, and inequality in India," *World Development*, 97, 102–121.

ARGENTE, D., AND M. LEE (Forthcoming): "Cost of living inequality during the great recession," *Journal of the European Economic Association.*

ATKIN, D., B. FABER, AND M. GONZALEZ-NAVARRO (2018): "Retail globalization and household welfare: Evidence from Mexico," *Journal of Political Economy*, 126(1), 1–73.

BANKS, J., R. BLUNDELL, AND A. LEWBEL (1997): "Quadratic Engel curves and consumer demand," *Review of Economics and Statistics*, 79(4), 527–539.

BILS, M., AND P. J. KLENOW (2001): "Quantifying quality growth," *American Economic Review*, 91(4), 1006–1030.

BLACKORBY, C., R. DAVIDSON, AND W. SCHWORM (1991): "Implicit separability: Characterisation and implications for consumer demands," *Journal of Economic Theory*, 55(2), 364–399.

BLACKORBY, C., D. PRIMONT, AND R. R. RUSSELL (1978): *Duality, separability, and functional structure: Theory and economic applications*, vol. 2. Elsevier Science Ltd.

BORUSYAK, K., AND X. JARAVEL (2018): "The distributional effects of trade: Theory and evidence from the United States," *Harvard University working paper.*

COMIN, D. A., D. LASHKARI, AND M. MESTIERI (2015): "Structural change with long-run income and price effects," *NBER working paper.*

CORNELSEN, L., R. GREEN, R. TURNER, A. D. DANGOUR, B. SHANKAR, M. MAZZOCCHI, AND R. D. SMITH (2015): "What happens to patterns of food consumption when food prices change? Evidence from a systematic review and meta-analysis of food price elasticities globally," *Health Economics*, 24(12), 1548–1559.

COSTA, D. L. (2001): "Estimating real income in the United States from 1888 to 1994: Correcting CPI bias using Engel curves," *Journal of Political Economy*, 109(6), 1288–1310.

DEATON, A. (2003a): "Adjusted Indian poverty estimates for 1999-2000," *Economic and Political Weekly*, pp. 322–326.

——— (2003b): "Prices and poverty in India, 1987-2000," *Economic and Political Weekly*, pp. 362–368.

DEATON, A., AND V. KOZEL (2005): "Data and dogma: The great Indian poverty debate," *World Bank Research Observer*, 20(2), 177–200.

DEATON, A., AND J. MUELLBAUER (1980): *Economics and Consumer Behavior*. Cambridge University Press.

DEATON, A., AND A. TAROZZI (2005): "Prices and Poverty in India," in *The Great Indian Poverty Debate*, ed. by A. Deaton, and V. Kozel, chap. 16. MacMillan.

FAJGELBAUM, P., AND A. KHANDELWAL (2016): "Measuring the unequal gains from trade," *The Quarterly Journal of Economics*, 131(3), 1113–1180.

FALLY, T. (2018): "Integrability and generalized separability," *CEPR discussion paper 12667*.

FEENSTRA, R. C. (1994): "New product varieties and the measurement of international prices," *American Economic Review*, 84(1), 155–57.

GOLDBERG, P. K., A. K. KHANDELWAL, N. PAVCNIK, AND P. TOPALOVA (2010): "Imported intermediate inputs and domestic product growth: Evidence from India," *Quarterly Journal of Economics*, 125(4), 1727–1767.

GORMAN, W. M. (1965): "Production functions in which the elasticities of substitution stand in fixed proportions to each other," *Review of Economic Studies*, 32(3), 217–224.

GORMAN, W. M. (1970): "Quasi-separable preferences, costs, and technologies," *University of North Carolina, Chapel Hill, mimeo*.

GORMAN, W. M. (1976): "Tricks with utility functions," *Essays in economic analysis: The proceedings of the Association of University Teachers of Economics, Sheffield 1975*.

GORMAN, W. M. (1995): *Collected works of W.M. Gorman: Separability and aggregation*, vol. 1. Oxford University Press.

HAMILTON, B. W. (2001): "Using Engel's Law to estimate CPI bias," *American Economic Review*, 91(3), 619–630.

HANOCH, G. (1975): "Production and demand models with direct or indirect implicit additivity," *Econometrica*, pp. 395–419.

HOTTMAN, C., AND R. MONARCH (2018): "Estimating unequal gains across U.S. consumers with supplier trade data," *Federal Reserve Board working paper*.

JARAVEL, X. (2018): "The unequal gains from product innovations: Evidence from the U.S. retail sector," *Quarterly Journal of Economics*, 134(2), 715–783.

JONES, C. I., AND P. J. KLENOW (2016): "Beyond GDP? Welfare across countries and time," *American Economic Review*, 106(9), 2426–57.

LEWBEL, A. (1991): "The rank of demand systems: Theory and nonparametric estimation," *Econometrica*, pp. 711–730.

LEWBEL, A., AND K. PENDAKUR (2009): "Tricks with Hicks: The EASI demand system," *American Economic Review*, 99(3), 827–863.

LIGON, E. (2019): "Estimating household welfare from disaggregate expenditures," *Working paper*.

MATSUYAMA, K. (2015): "The home market effect and patterns of trade between rich and poor countries," *Northwestern University manuscript*.

MCFADDEN, D. (1978): "Cost, revenue, and profit functions," in *Production economics: A dual approach to theory and applications*, ed. by M. U. A. for the History of Economic Thought, vol. 1.

MUKHERJEE, A. (2015): "Services sector in India: trends, issues, and the way forward," *Eurasian Geography and Economics*, 56(6), 635–655.

NAKAMURA, E., J. STEINSSON, AND M. LIU (2016): "Are Chinese growth and inflation too smooth? Evidence from Engel curves," *American Economic Journal: Macroeconomics*, 8(3), 113–44.

NAKAMURA, L. (1996): "Is U.S. economic performance really that bad?," *Federal Reserve Bank of Philadelphia Working Paper No. 95-21/R.*

REDDING, S. J., AND D. E. WEINSTEIN (2020): "Measuring aggregate price indices with taste shocks: Theory and evidence for CES preferences," *The Quarterly Journal of Economics*, 135(1), 503–560.

SARHAN, A. (1955): "Estimation of the mean and standard deviation by order statistics. Part II," *The Annals of Mathematical Statistics*, 26(3), 505–511.

TAROZZI, A. (2007): "Calculating comparable statistics from incomparable surveys, with an application to poverty in India," *Journal of Business & Economic Statistics*, 25(3), 314–336.

TOPALOVA, P. (2010): "Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India," *American Economic Journal: Applied Economics*, 2(4), 1–41.

WORKING, H. (1943): "Statistical laws of family expenditures," *Journal of the American Statistical Association*, 38, 43–56.

YOUNG, A. (2012): "The African growth miracle," *Journal of Political Economy*, 120(4), 696–739.

8 Figures and Tables

8.1 Figures

Figure 1: Shifts in Engel Curves Over Time and Across Space
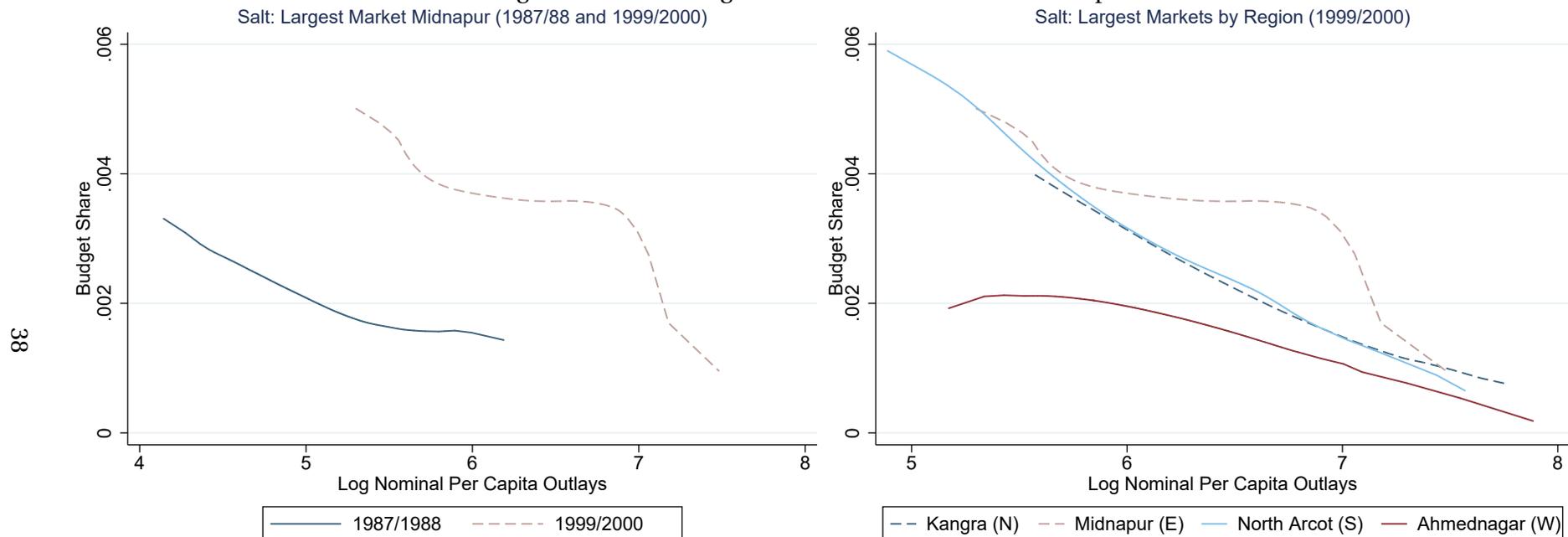
Notes: Figures plot Engel curves for salt over time (NSS 43rd Round 1987-1988 to NSS 55th round 1999-2000) for the largest rural market (Midnapur), and over space for the largest markets in the four broad regions of India (in terms of numbers of households surveyed). A market is defined as the rural area of an Indian district. Fitted relationships are based on local polynomial regressions using an Epanechnikov kernel with the "rule-of-thumb" bandwidth estimator. See Section 3 for further discussion.

## Figure 2: Illustration of Lemma 1
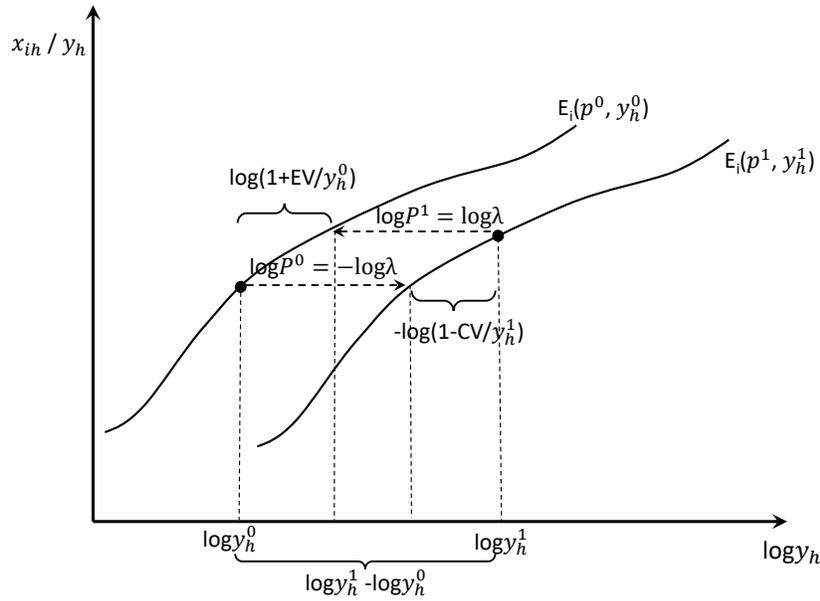
Notes: Figure illustrates how price indices and welfare can be recovered from horizontal shifts in Engel curves (i.e. budget shares plotted against log total expenditure per capita) when all relative prices are unchanged. Period 0 and period 1 Engel curves for good i denoted by $E_i(p^0, y_h^0)$ and $E_i(p^1, y_h^1)$, respectively. See Section 4 for further discussion.

Figure 3: Illustration of Proposition 1
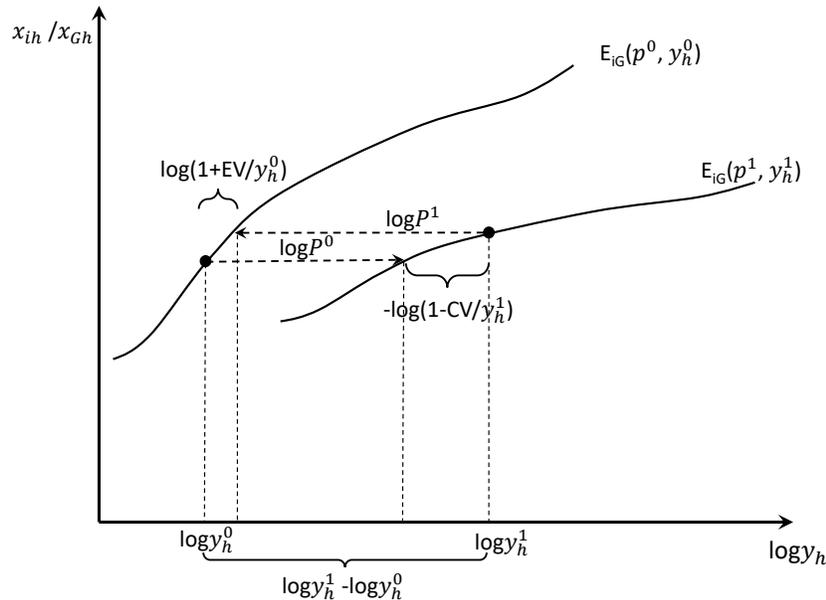
Notes: Figure illustrates how price indices and welfare can be recovered from horizontal shifts in relative Engel curves (i.e. expenditure on good i as a share of total expenditure on group G plotted against log total expenditure per capita) when relative prices within group G are unchanged but prices outside of G are unrestricted. Period 0 and period 1 relative Engel curves for good i denoted by $E_{iG}(p^0, y_h^0)$ and $E_{iG}(p^1, y_h^1)$, respectively. See Section 4 for further discussion.

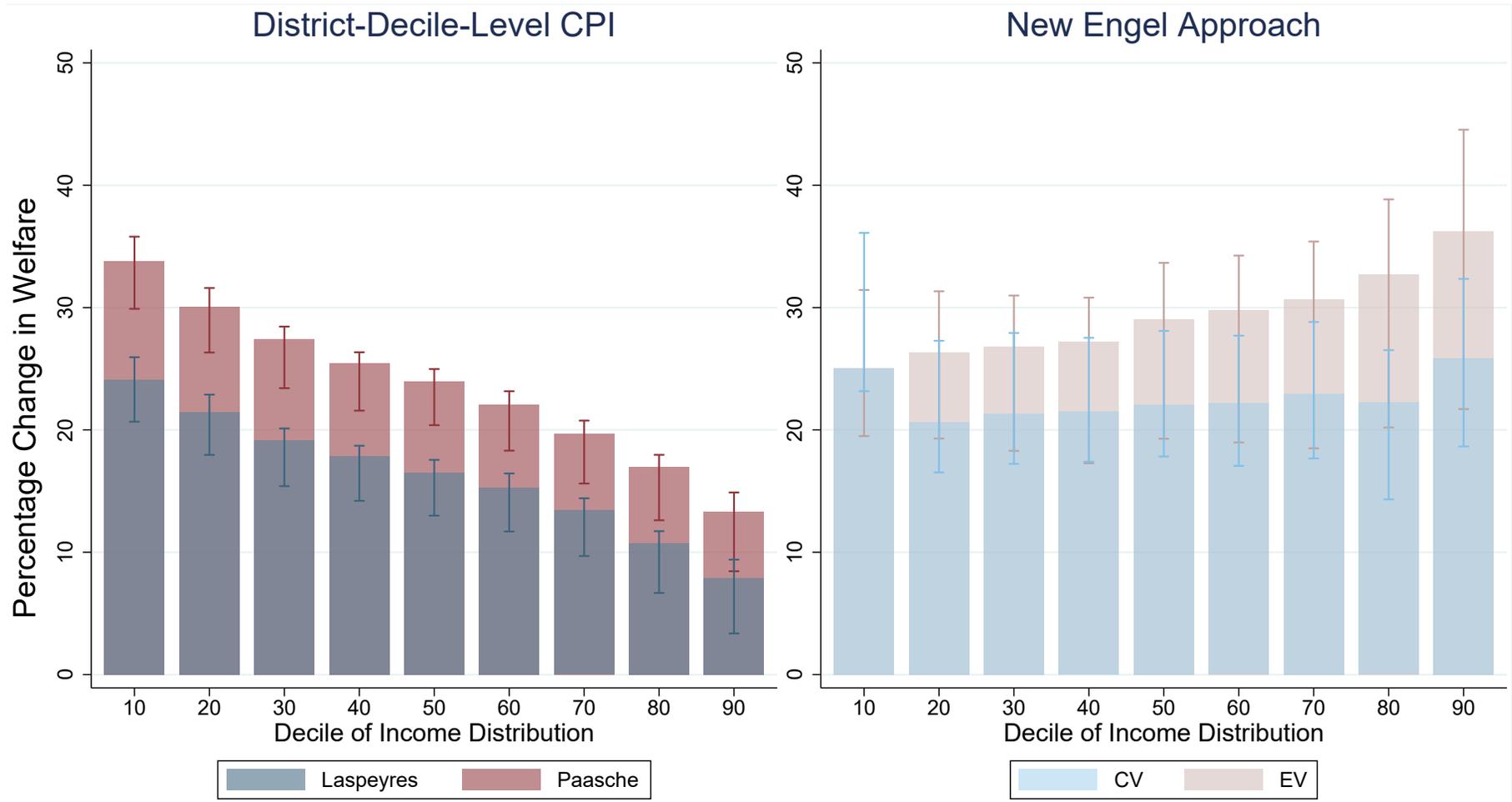Figure 4: Rural Indian Growth in Nominal Outlays Per Capita 1987/88-1999/2000
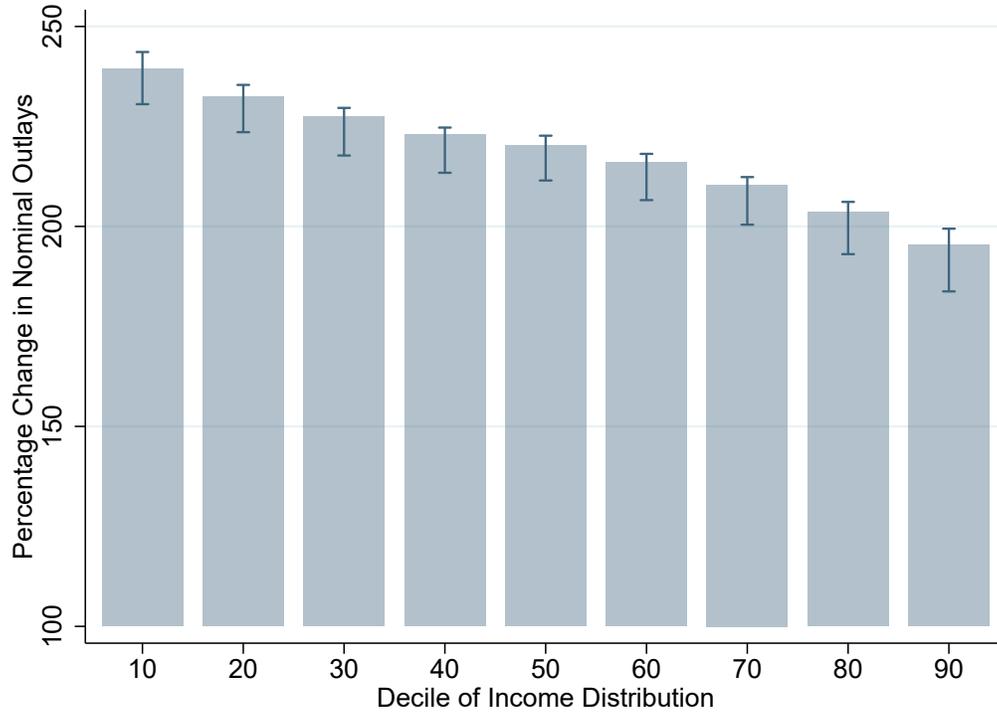
Notes: Figure shows the percentage change in rural nominal outlays per capita between 1987/88 and 1999/2000 for each decile of the local nominal outlay distribution (averaged across districts using population weights). Bootstrapped confidence intervals are based on sampling with replacement 1000 times from the distribution of households within each district-survey round and plotting the 2.5 and 97.5 percent envelope of price index estimates at each decile. See Section 6.2 for further discussion.

Figure 5: Rural Indian Cost of Living Inflation 1987/88-1999/2000: New Engel Approach Compared to Existing CPI Estimates
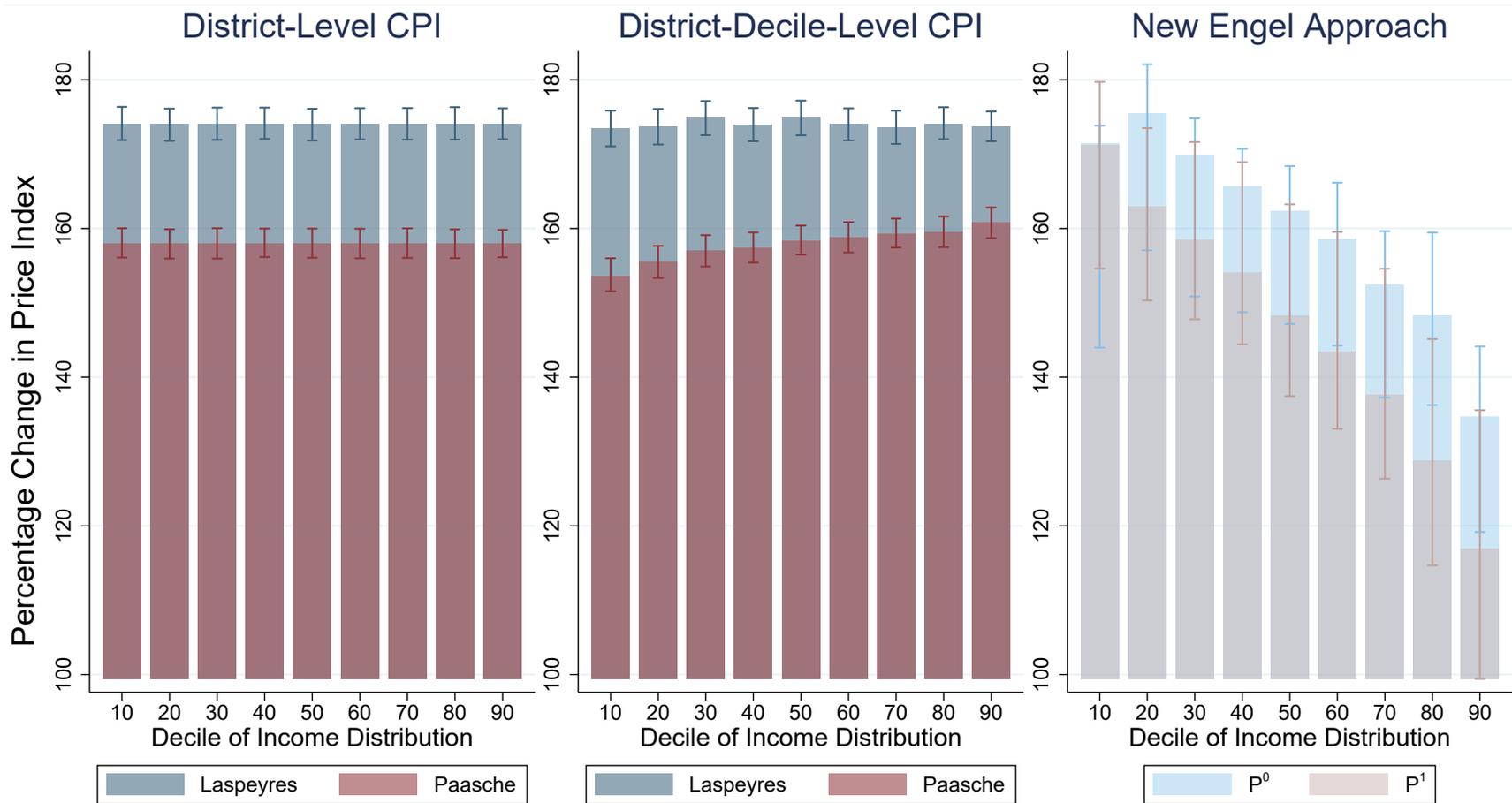
Notes: Figure shows the percentage change in the rural price index between 1987/88 and 1999/2000 for each decile of the local nominal outlay distribution (averaged across districts using population weights). Left panel plots price index changes using Laspeyres and Paasche district-level CPIs calculated using price changes for food and fuels following Deaton (2003b). Middle panel repeats the left panel but using district-income-decile-specific budget shares to calculate the Laspeyres and Paasche indices. Right panel plots our New Engel Approach price index changes estimated from horizontal shifts in relative Engel curves. Bootstrapped confidence intervals are based on sampling with replacement 1000 times from the distribution of households within each district-survey round and plotting the 2.5 and 97.5 percent envelope of price index estimates at each decile. See Section 6.2 for further discussion.

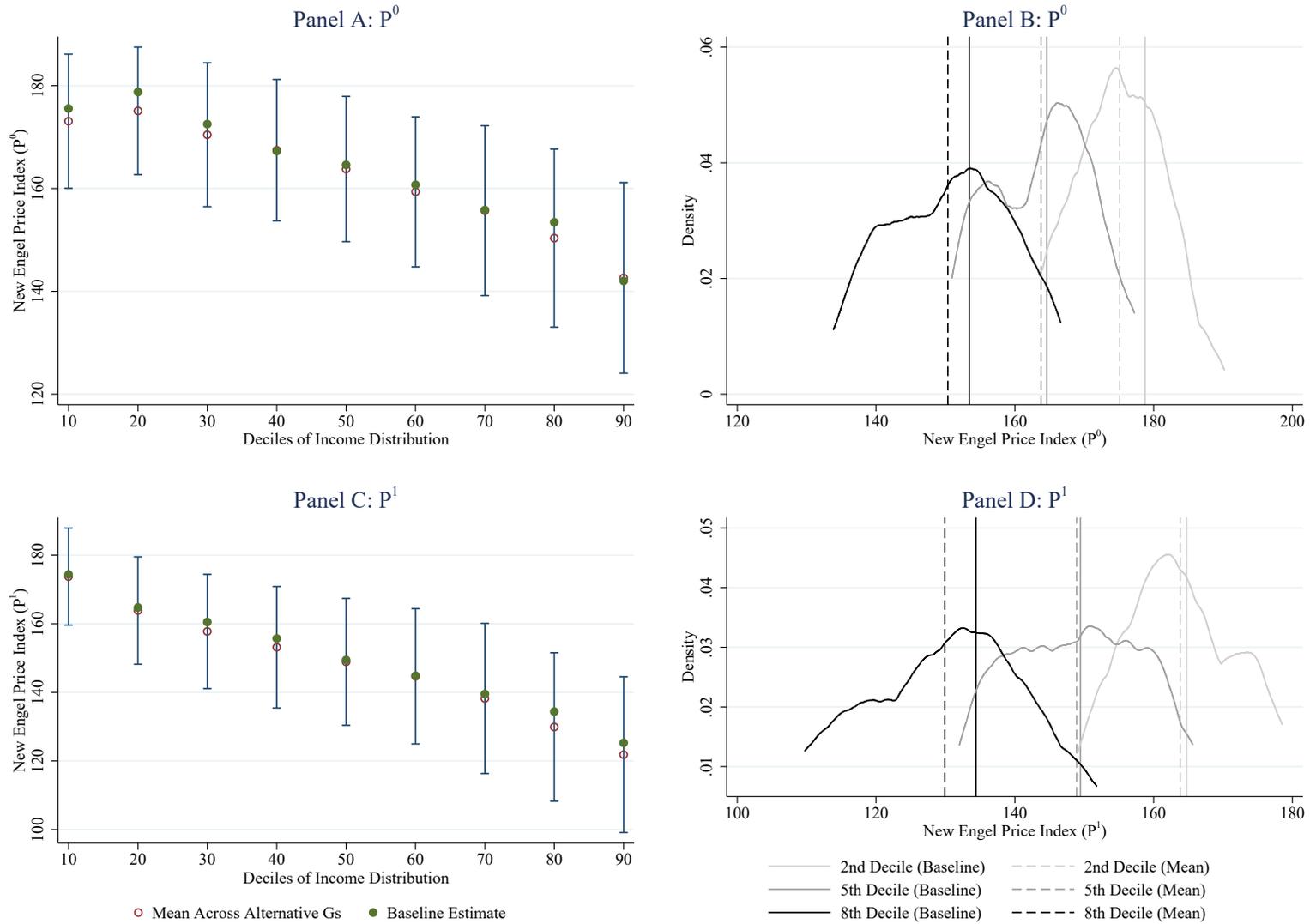Figure 6: Rural Indian Welfare Growth 1987/88-1999/2000

District-Decile-Level CPI

New Engel Approach

Legend: Laspeyres, Paasche

Legend: CV, EV

Notes: Figure shows the percentage change in rural welfare between 1987/88 and 1999/2000 for each decile of the local nominal outlay distribution (averaged across districts using population weights). Left panel plots real income calculated by deflating nominal income changes by Laspeyres and Paasche price index changes (using price changes for food and fuels and district-income-decile-specific budget shares). Percentage change computed as $100((y_h^1/y_h^0)/P_h - 1)$ where P_h is the proportional price index change. Right panel plots New Engel Approach welfare changes estimated from nominal income changes and horizontal shifts in relative Engel curves. Percentage changes using equivalent variation are computed as $100(EV_h/y_h^0)$ (or, equivalently, $100((y_h^1/y_h^0)/P^1 - 1)$ where $P^1 = P^1(y_h)$ is our New Engel Price Index holding utility at period 1's level). Percentage changes using compensating variation are computed as $100(CV_h/(y_h^1 - CV_h))$ where the denominator $y_h^1 - CV_h$ is base period nominal incomes under period 1 prices (equivalently, $100((y_h^1/y_h^0)P^0 - 1)$ where $P^0 = P^0(y_h)$ is our New Engel Price Index holding utility at period 0's level). Bootstrapped confidence intervals are based on sampling with replacement 1000 times from the distribution of households within each district-survey round and plotting the 2.5 and 97.5 percent envelope of price index estimates at each decile. See Section 6.2 for further discussion.

Figure 7: Rural Indian Cost of Living Inflation 1987/88-1999/2000: First-Order and Exact Price Corrections
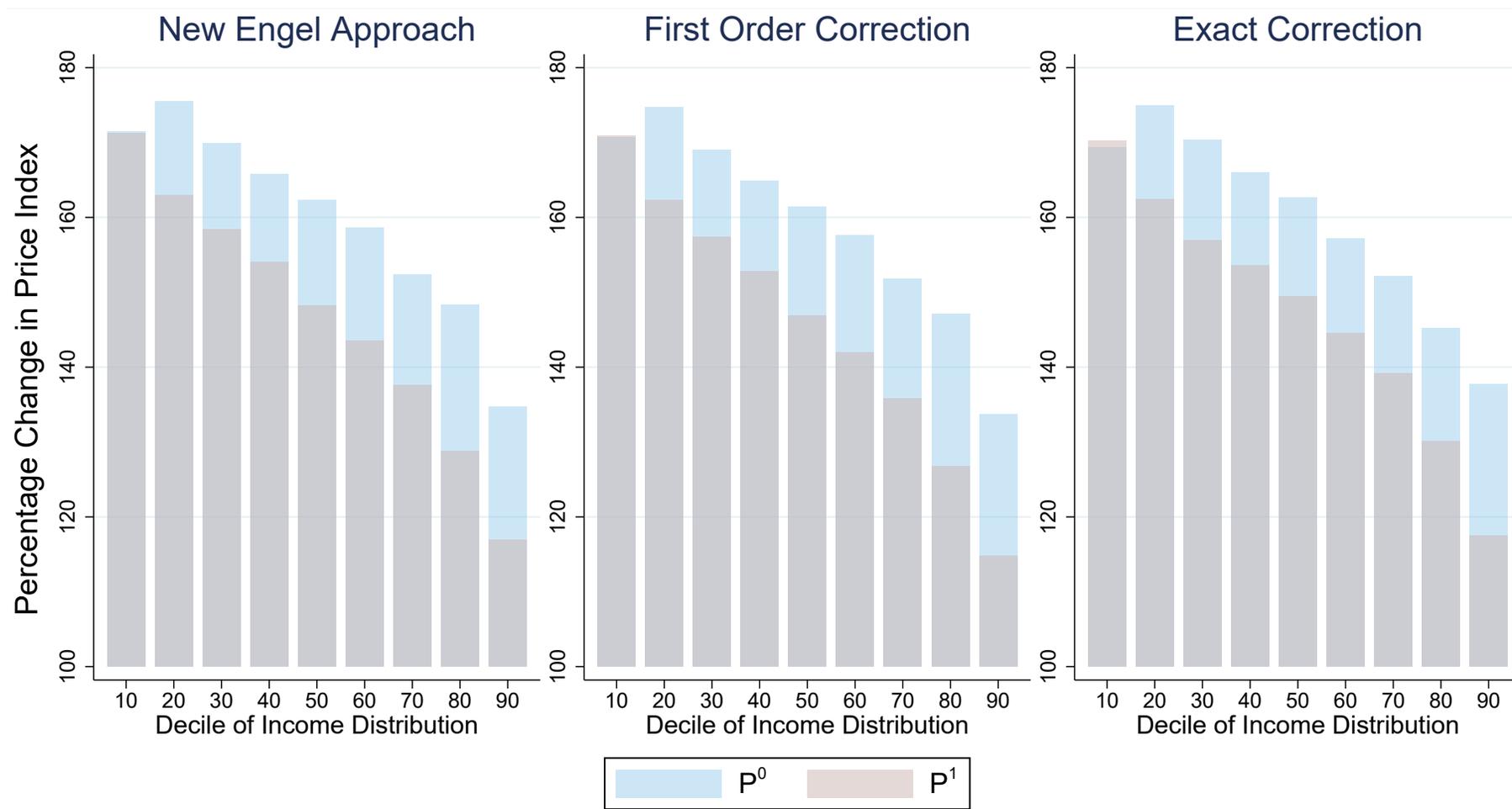
*Notes*: Figure shows corrected New Engel price index changes to account for relative price changes within-G groups. Left panel repeats the uncorrected price index change shown in Figure 5. Middle panel applies the first-order correction and right panel applies the exact correction, both described in Section 5.2.2, using an elasticity of substitution of 0.7. See Section 6.2.1 for further discussion.

Figure 8: New Engel Price Index Changes Across Alternative G Groups

Panel A: P^0

Panel B: P^0

Panel C: P^1

Panel D: P^1

○ Mean Across Alternative Gs   ● Baseline Estimate

2nd Decile (Baseline) ------ 2nd Decile (Mean)
5th Decile (Baseline) ------ 5th Decile (Mean)
8th Decile (Baseline) ------ 8th Decile (Mean)

Notes: Figure reports New Engel rural price index changes by local income decile (averaged across districts using population weights) for each of 108 alternative classifications of goods into plausibly quasi-separable groups G. Our baseline classification of three quasi-separable groups is one of the 108 classifications, and we indicate our baseline estimates in all panels. The two left panels depict the mean and 95% confidence intervals across the 108 alternative estimates for each decile (panel A for P^0 and panel B for P^1). The two right panels depict the distribution of these estimates for the 2nd, 5th and 8th deciles of local nominal incomes (panel B for P^0 and panel D for P^1). See Section 6.2.1 for further discussion.

Figure 9: Effect of Import Competition on Rural Nominal Outlays Per Capita and Welfare
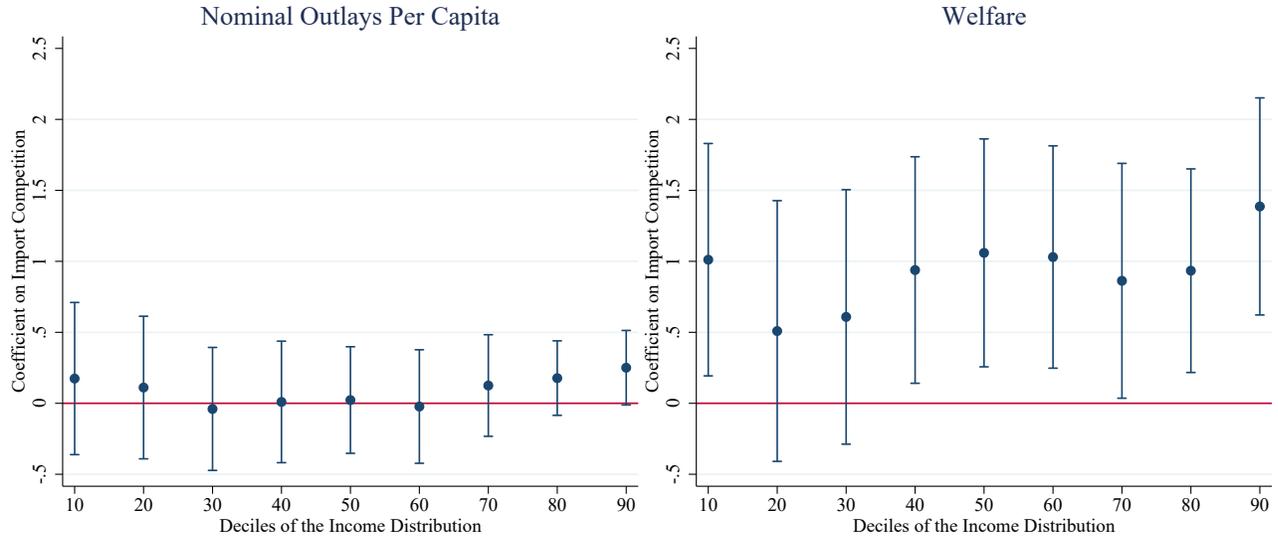
Notes: The left panel depicts IV point estimates of the effect of import competition on log nominal outlays per capita, estimated separately for each decile of the local income distribution. The IV regression specification follows column 8 in Table 3a) of Topalova (2010). Specifically, exposure to import competition is measured by the weighted average tariff cut, with weights proportional to the initial sectoral employment shares in the district. There are two instruments: first the same shift-share measure but calculated only using tradable industries, second this tradable shift-share but using the initial average level of import tariffs rather than the change. Specification also includes district fixed effects, time fixed effects, and additional time-varying district controls. The right panel depicts estimates from identical specifications with log welfare (measured by equivalent variation) as the dependent variable. 95 percent confidence intervals based on standard errors clustered at the state-by-survey-round level (as in Topalova). See Section 6.3 for further discussion.

8.2 Tables

Table 1: Two Motivating Facts about Engel Curves in Rural India

	Motivating Fact 1 Fraction of Goods for which Linear Engel Curves Rejected	Motivating Fact 2 Fraction of Good-by-Market Cells for which Uniform Shift in Engel Curves Across Periods Rejected
Fraction Rejecting at the 99% Confidence Level	0.815	0.567
Fraction Rejecting at the 95% Confidence Level	0.899	0.686
Fraction Rejecting at the 90% Confidence Level	0.905	0.745

Notes: Table presents two motivating facts about Engel curves in our rural Indian sample. Engel curves in the first column are estimated separately by good by stacking within-market-by-period variation across all markets and both periods. We reject linearity if the joint test of all 2nd or higher-order polynomial terms (up to the 4th order) of log household total outlay per capita are significantly different from zero. The second column uses good-by-market-by-period estimates of Engel curves (two for each good and market) covering all markets with at least 100 households in both survey rounds. We reject a uniform horizontal shift in an Engel curve across periods if the shift in log nominal outlays per capita, moving from Round 43 to 55, is not uniform for different levels of budget shares (i.e. if the interactions between a Round 55 dummy and 4th order polynomial terms of the budget share are not jointly significant at the indicated levels). All regressions are weighted by survey weights and standard errors are clustered at the market (district) level. See Section 3 for further discussion.

Table 2: Quasi-Separability Test

Panel A	Dep. var: log New Engel Price Index (calculated excluding fuels)			
	log P^0	log P^0	log P^1	log P^1
Δlog(Price Fuel)*Exp. Share Fuel	1.136***	1.115***	0.928***	0.877***
	(0.180)	(0.180)	(0.196)	(0.199)
p-value test β=1	0.45	0.52	0.71	0.54
Obs	2926	2926	2986	2986
R2	0.0350	0.0350	0.0316	0.0310
Decile specific Δlog(Price Fuel)	No	Yes	No	Yes
Income decile FE	Yes	Yes	Yes	Yes

| Panel B	Dep. var: log New Engel Price Index (exact correction, calculated excluding fuels)			
	log P^0	log P^0	log P^1	log P^1
Δlog(Price Fuel)*Exp. Share Fuel	1.011***	0.981***	0.913***	0.867***
	(0.174)	(0.173)	(0.189)	(0.191)
p-value test β=1	0.95	0.91	0.65	0.49
Obs	2934	2934	2986	2986
R2	0.0215	0.0212	0.0262	0.0256
Decile specific Δlog(Price Fuel)	No	Yes	No	Yes
Income decile FE	Yes	Yes	Yes	Yes

Notes: Table performs the quasi-separability test described in subsection 5.2.4. Dependent variable in Panel A is the log New Engel price index, either $\log P^0$ or $\log P^1$, estimated using only food items (i.e. excluding fuels). In Panel B the dependent variable is the log New Engel price index estimated using only food items (excluding fuels) correcting for relative price changes using the isoleastic correction (non-homothetic CES) in equation (10) with $\sigma = 0.7$. The explanatory variable in columns 1 and 3 is the log change in the price of fuels (calculated using a Paasche price index of fuel items where weights are given by mean district-level expenditure shares across items within the fuels category) multiplied by the district-by-decile expenditure share on fuels. The explanatory variable in columns 2 and 4 uses the decile-specific log change in the price of fuels (calculated using a Paasche price index of fuel items where weights are given by district-by-decile mean expenditure shares across items within the fuels category) multiplied by the district-by-decile expenditure share on fuels. The first row of the bottom panel reports the *p*-value on the test of the coefficient of interest being equal to 1, as required by the quasi-separability test. Regressions are weighted using district weights. Robust standard errors in parenthesis.

Appendix

A Additional Figures and Tables

Appendix Figures

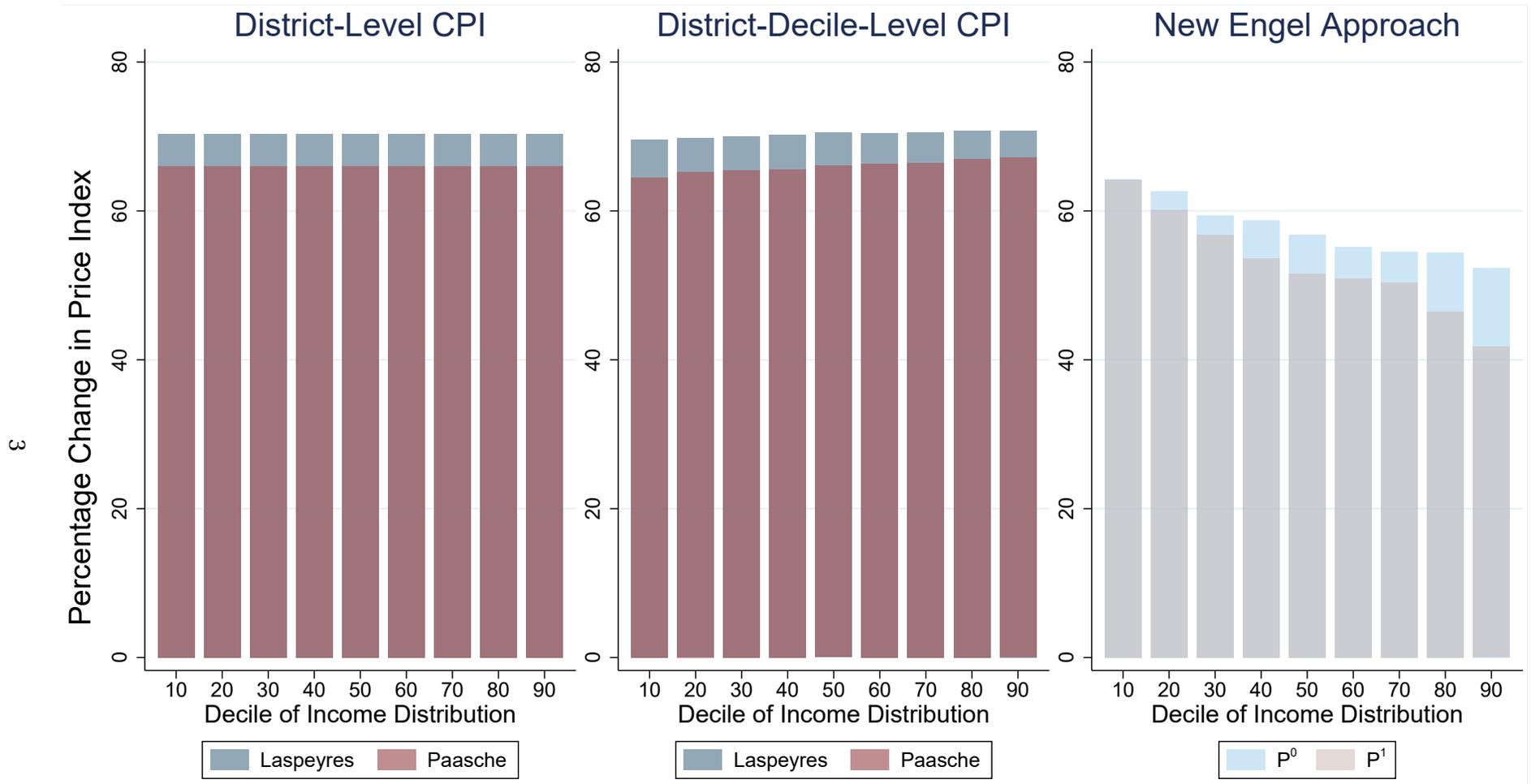Figure A.1: Sparseness Across Alternative Product Aggregations
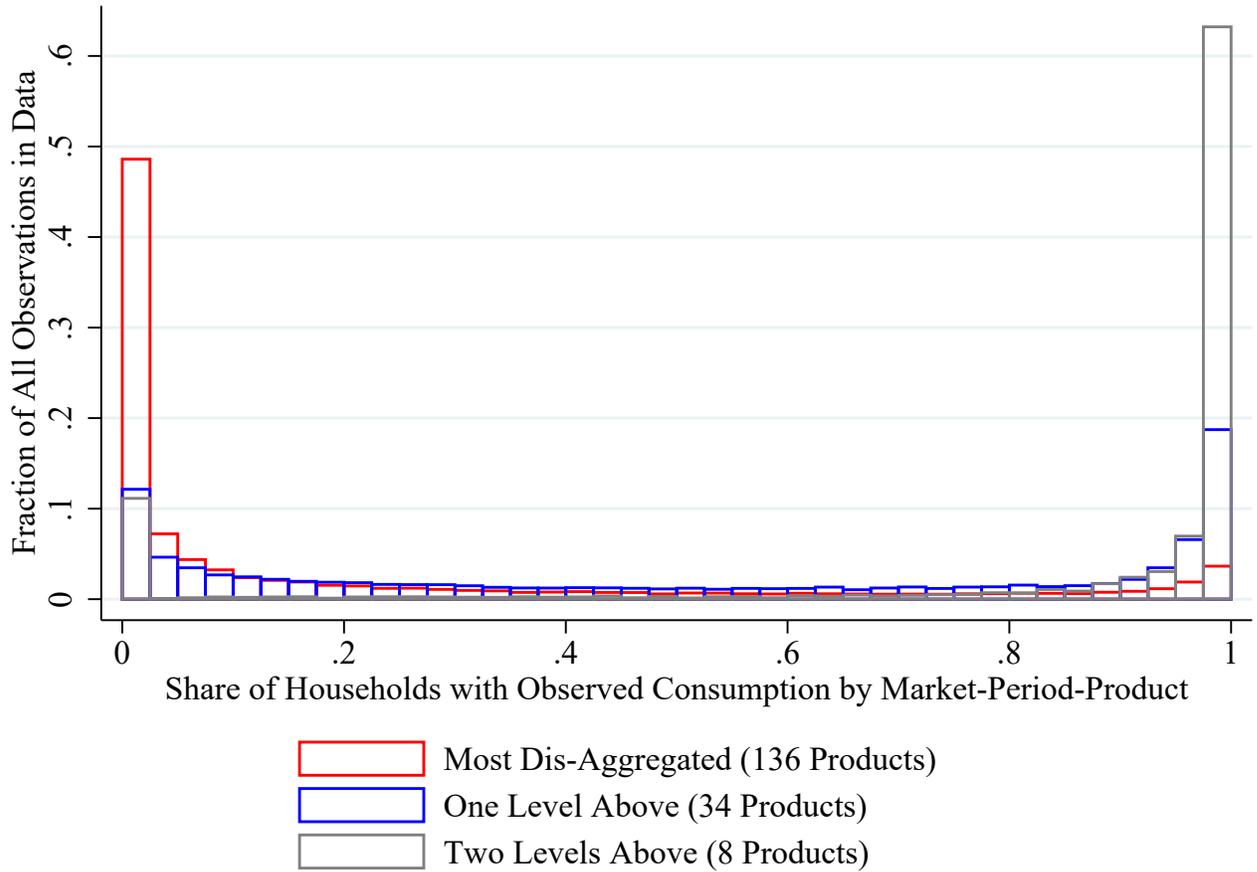
Notes: Figure plots histogram of share of households with any observed consumption by product-period-market cell across three alternative levels of product aggregation. See Section 6.1 for further discussion.

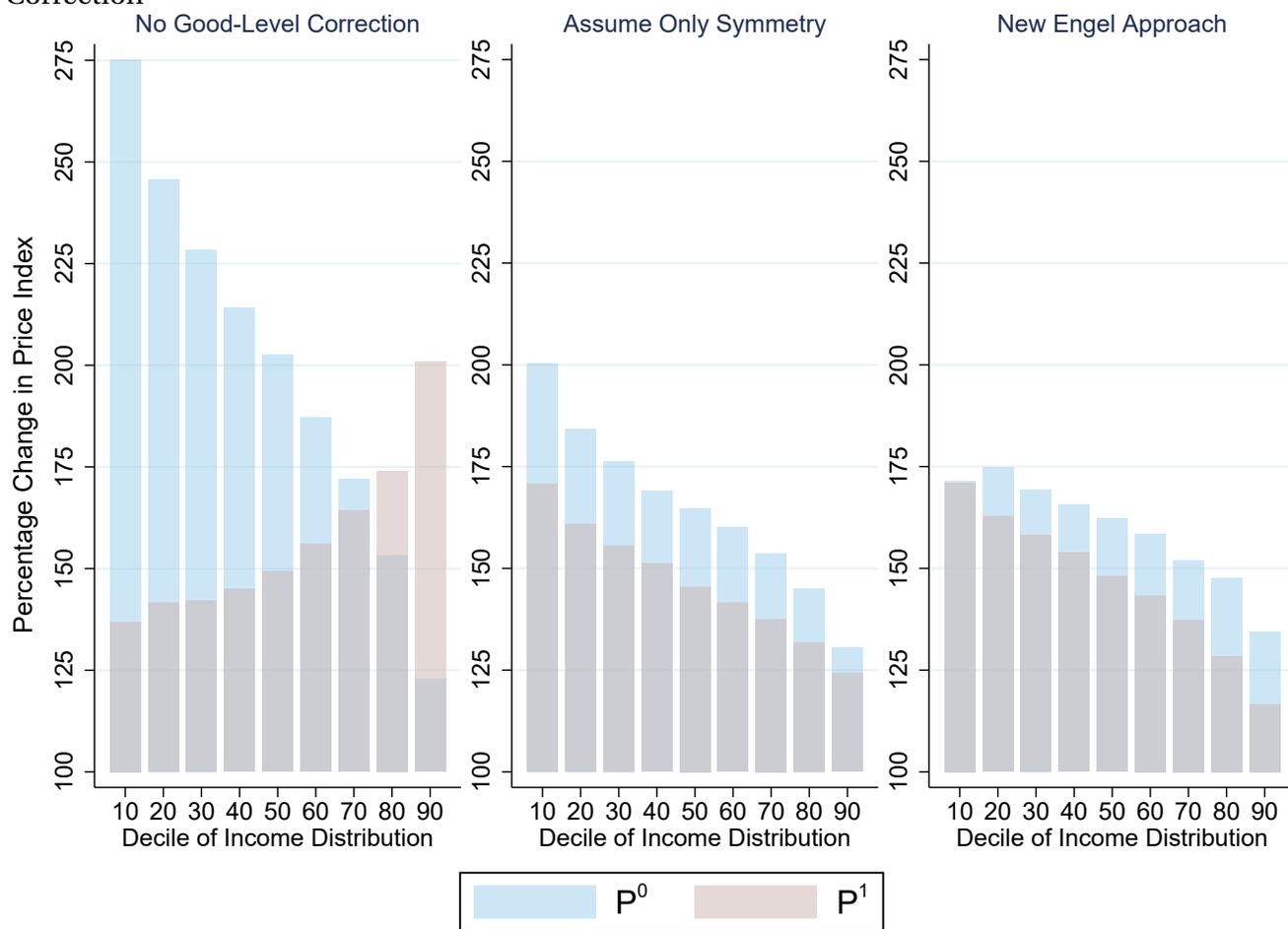Figure A.2: Price Index Estimates Using Alternative Levels of Good Aggregation
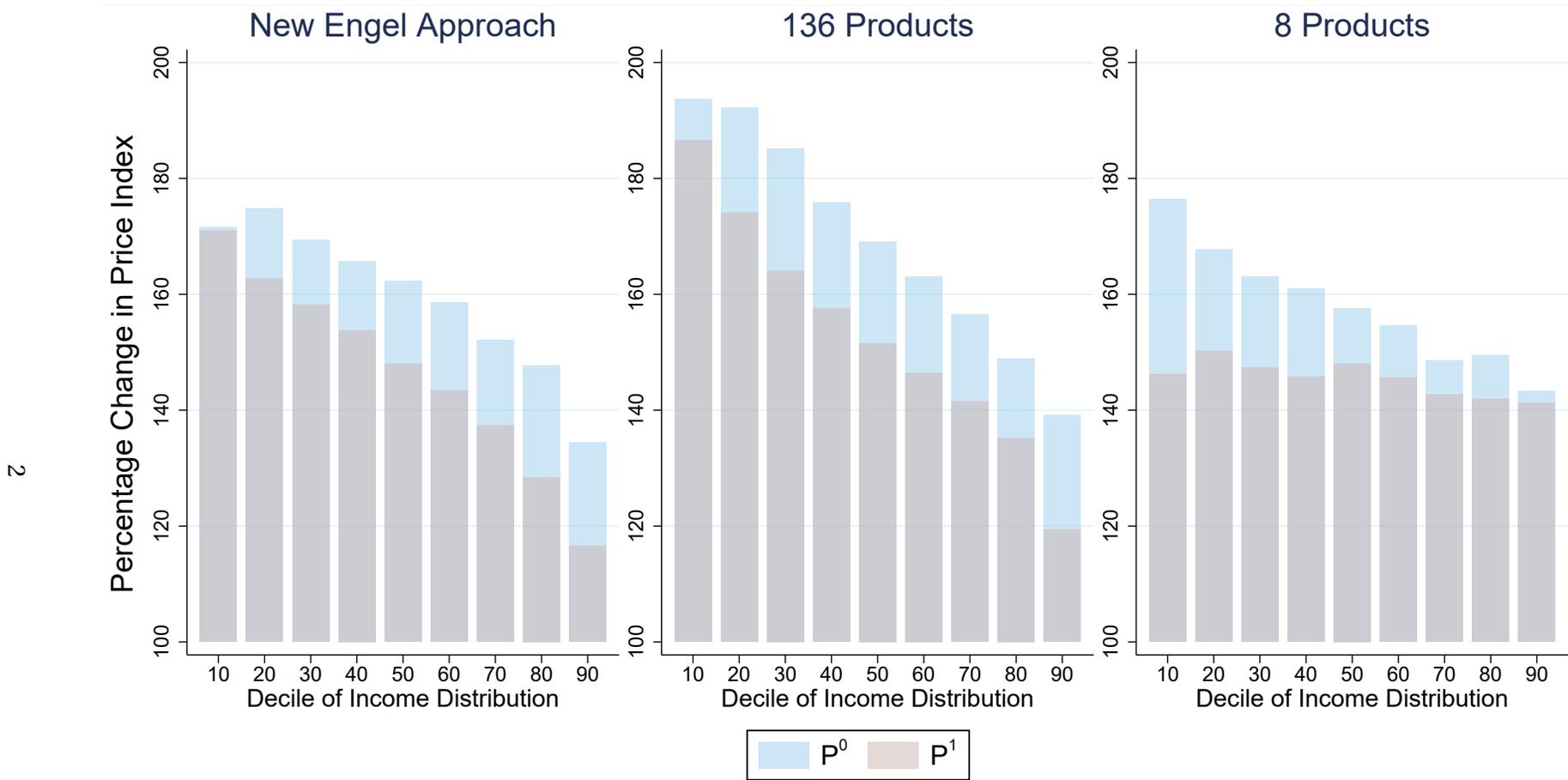
Notes: Figure shows the average percentage change in the rural price index between 1987/88 and 1999/2000 for each decile of the local nominal outlay distribution (averaged across districts using population weights). Estimates are based on New Engel Approach using horizontal shifts in relative Engel curves. The three panels use different levels of aggregation of goods in the Indian expenditure microdata. The left panel depicts our baseline estimation approach which aggregates the 136 products to 34 products (the second-lowest level of aggregation in the NSS surveys). The middle panel uses the disaggregated 136 products, while the right panel further aggregates to 8 products (the third-lowest level of aggregation in the NSS surveys). See Section 6.1 for further discussion.

Figure A.3: Recall Bias: Rural Inflation 1987/88-1994/95

District-Level CPI | **District-Decile-Level CPI** | **New Engel Approach**

Laspeyres · Paasche (left and middle panels)
P^0 · P^1 (right panel)

Notes: Figure shows the percentage change in the rural price index between 1987/88 and 1994/1995 for each decile of the local nominal outlay distribution (averaged across districts using population weights). Left panel plots price index changes using Laspeyres and Paasche district-level CPIs calculated using price changes of food and fuels following Deaton (2003b). Middle panel repeats the left panel but using district-income-decile-specific budget shares to calculate the Laspeyres and Paasche indices. Right panel plots our New Engel Approach price index changes estimated from horizontal shifts in relative Engel curves. See Section 6.2 for further discussion.

3

Figure A.4: Good-Level Selection Corrections (1): Price Index Changes With and Without Bias Correction


Notes: Figure shows the percentage change in the rural New Engel price index between 1987/88 and 1999/2000 for each decile of the local nominal outlay distribution (averaged across districts using population weights), both with and without correcting estimates for selection bias described in Section 5.2.5. Left panel plots New Engel estimates that are simple averages of all overlapping Engel curves for a particular market. Middle panel accounts for bias from non-overlapping Engel curves by assuming distribution of price index estimates within a market is symmetric, ordering both overlapping and non-overlapping estimates, and taking the median when observed. The Right panel, our baseline New Engel Approach, further assumes the distribution is uniform to calculate medians when not observed. See Section 6.2.1 for further discussion.

Figure A.5: Good-Level Selection Corrections (2): Reasons for Non-Overlap
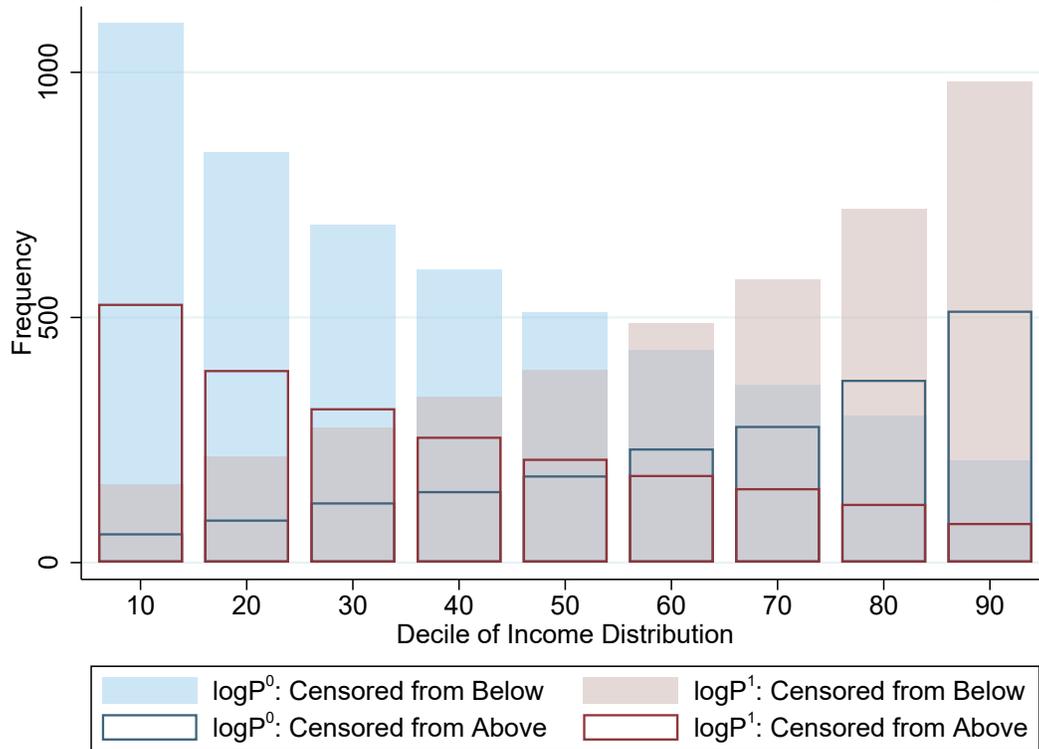
Notes: Figure shows the frequency of non-overlapping estimates by decile, broken out by type of non-overlap (censored from above or from below). This information is used to rank missing (non-overlapping) estimates and calculate the medians required for the good-level selection correction applied in both the middle and right panel of Appendix Figure A.4. See Section 6.2.1 for further discussion.

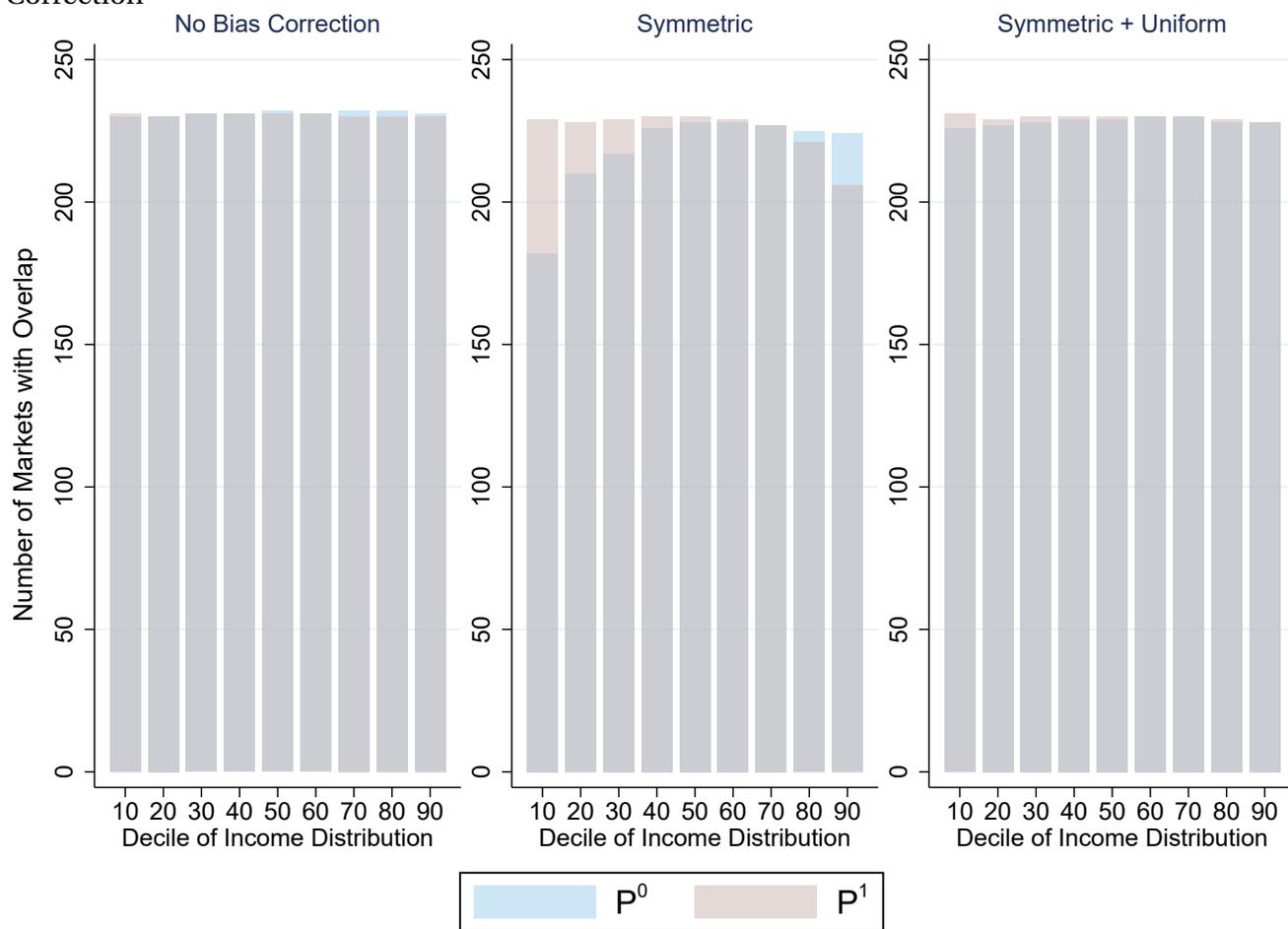Figure A.6: Good-Level Selection Corrections (3): Number of Markets With and Without Bias Correction

Notes: Figure shows the number of missing market-decile pairs after applying the good-level selection correction just using symmetry (middle panel) and symmetry plus uniformity (our baseline, right panel). For comparison, left panel shows the number of market-decile pairs where we have at least one good with overlapping monotonic relative Engel curves at that decile of the income distribution and so can obtain an estimate of the price index without any bias correction. See Section 6.2.1 for further discussion.

Figure A.7: New Engel Results After Including Household Controls
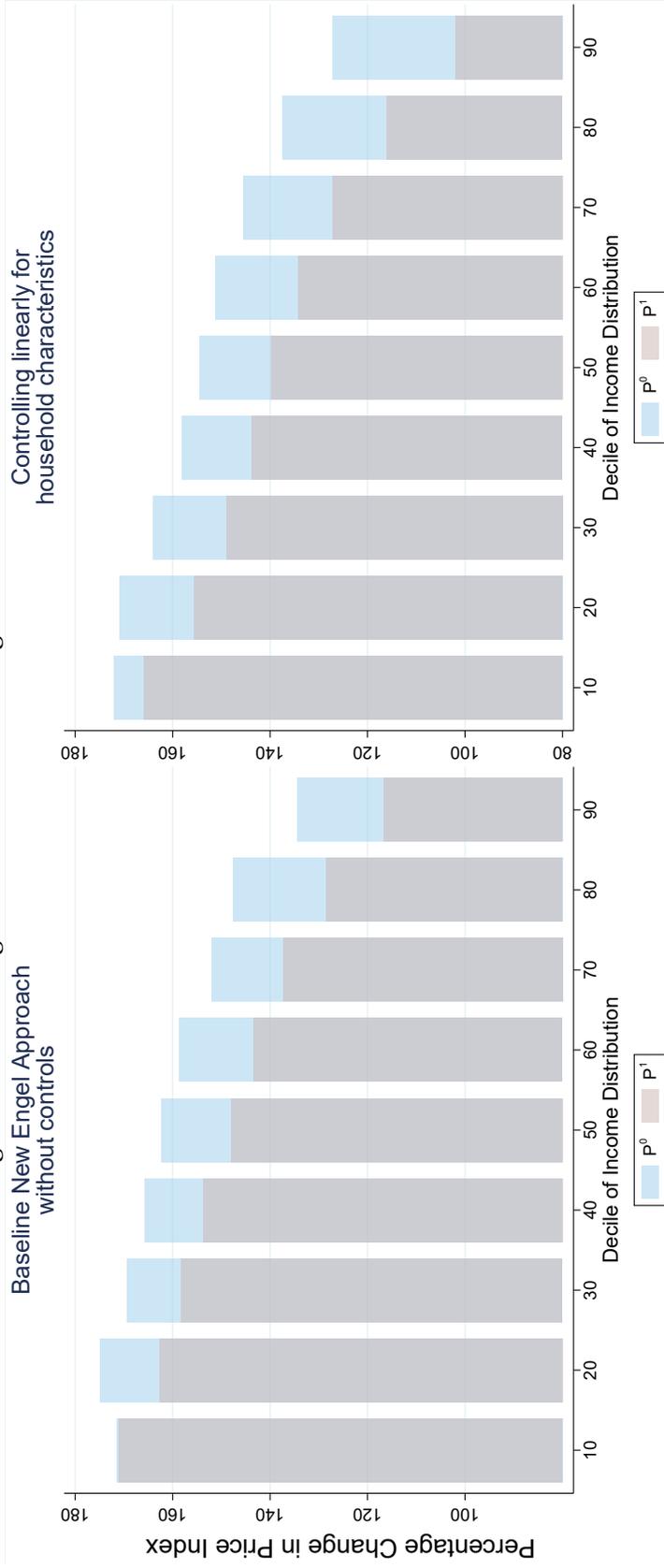
Notes: Left panel shows the baseline New Engel Approach price index estimates. Right panel shows New Engel estimates after conditioning on household controls when estimating relative Engel curves. Specifically, for each good and market, we regress relative budget shares against log income (non-parametrically), including linear controls for household characteristics (a scheduled caste dummy, a literacy of household head dummy, log of household size, and the share of children in the household). Coefficients from these market-good specific linear controls are used to evaluate relative budget shares at the market median value (constant over time) for each characteristic. We then use these characteristic-adjusted budget shares to obtain the New Engel price index changes shown in the right panel. See Section 6.2.1 for further discussion.

Figure A.8: New Engel Price Index Changes For Separate Groups of Households
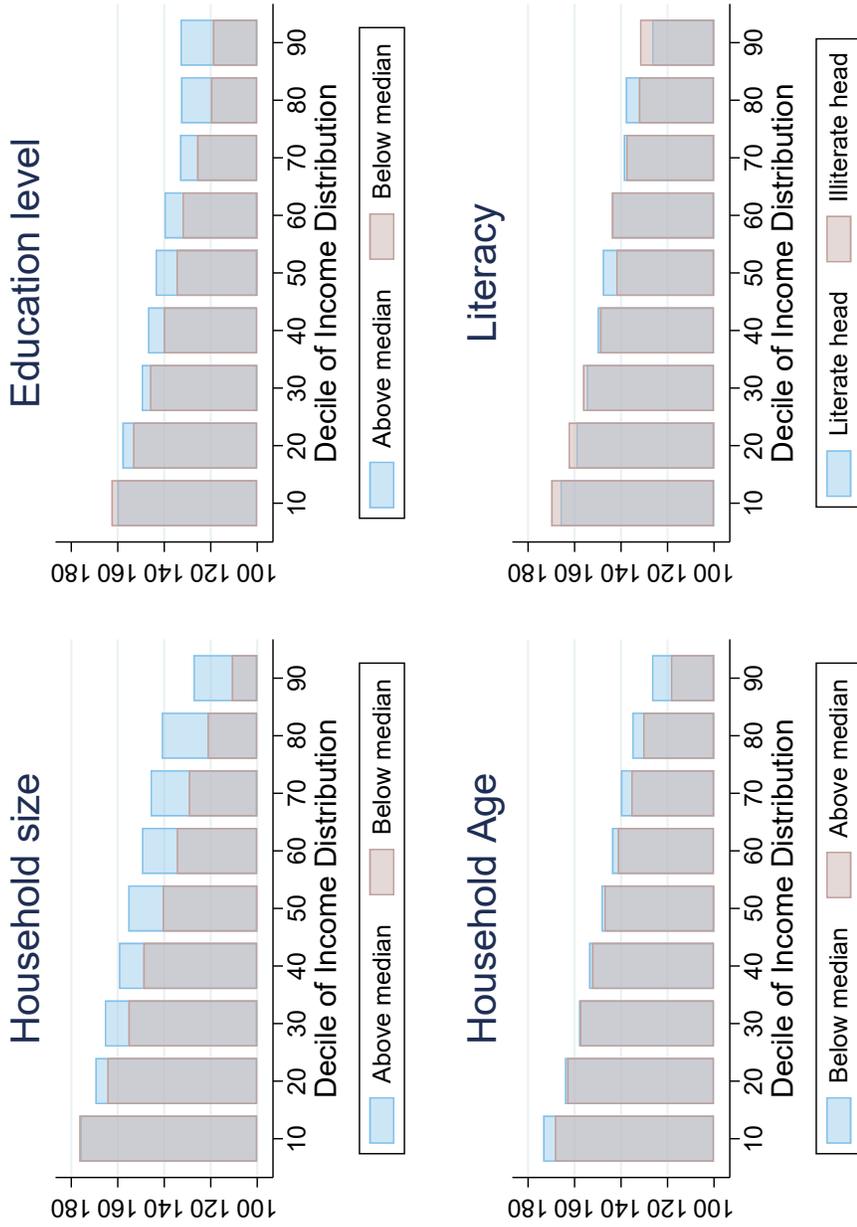
Notes: Figure shows New Engel rural price index changes by local income decile (averaged across districts using population weights) partitioning households within each market along four dimensions: (top-left) above and below median household size in the district, (top-right) above and below median household head education level in the district, (bottom-left) above and below median household head age in the district, (bottom-right) by literate/illiterate status of the household head.

Figure A.9: Alternative Estimates of Relative Engel Curves
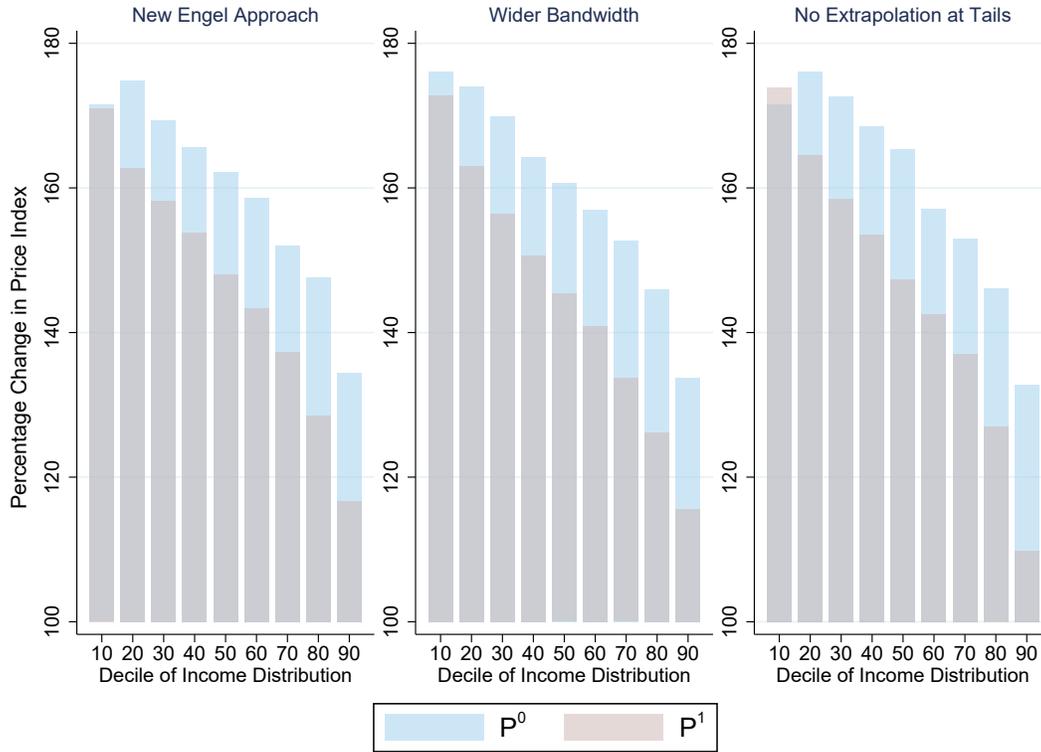
Notes: Figure shows New Engel price index changes using alternate methods of estimating relative Engel curves. Left panel reproduces our baseline approach. Recall that the baseline approach uses an Epanechnikov kernel for non-parametrically estimating Engel curves equal to one quarter of the range of the income distribution. Additionally, we restrict attention to good-market combinations where Engel curves in both periods are monotonic between percentiles 5 and 95 and drop relative expenditure share estimates beyond those percentiles in cases where those portions are non-monontonic—replacing those values with a linear extrapolation from the monotonic portion of the curve. Middle panel extends the bandwidth of the Epanechnikov kernel used to 30 percent of the range. Right panel does not replace extreme non-monotonic values with linear extrapolations.

Figure A.10: Rural Indian Cost of Living Inflation 1987/88-1999/2000: Using All Markets (Including Markets <100 Hholds)
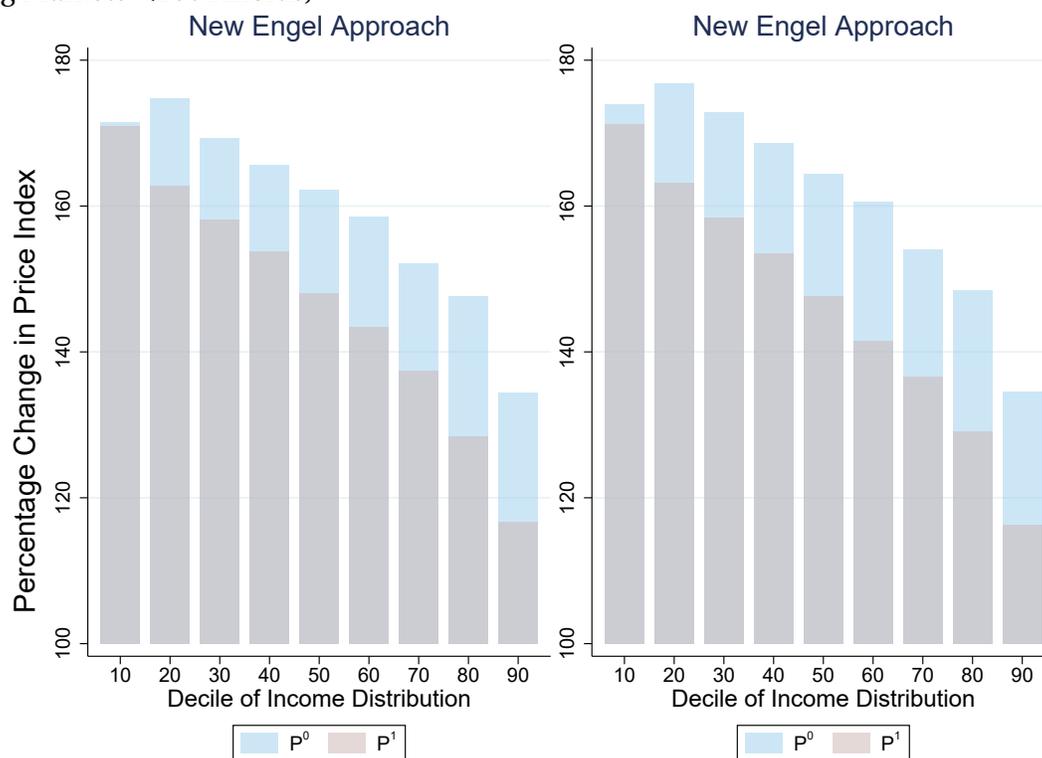
Notes: Figure shows the percentage change in the rural price index between 1987/88 and 1999/2000 for each decile of the local nominal outlay distribution (averaged across districts using population weights). Left panel plots our baseline New Engel price index changes that exclude small markets (those with fewer than 100 households surveyed in each survey round). Right panel plots New Engel price index changes including all markets.

Figure A.11: Effect of Import Competition on Laspeyres Price Index (Only Using Reliable Price Data)

District-Decile-Level Laspeyres
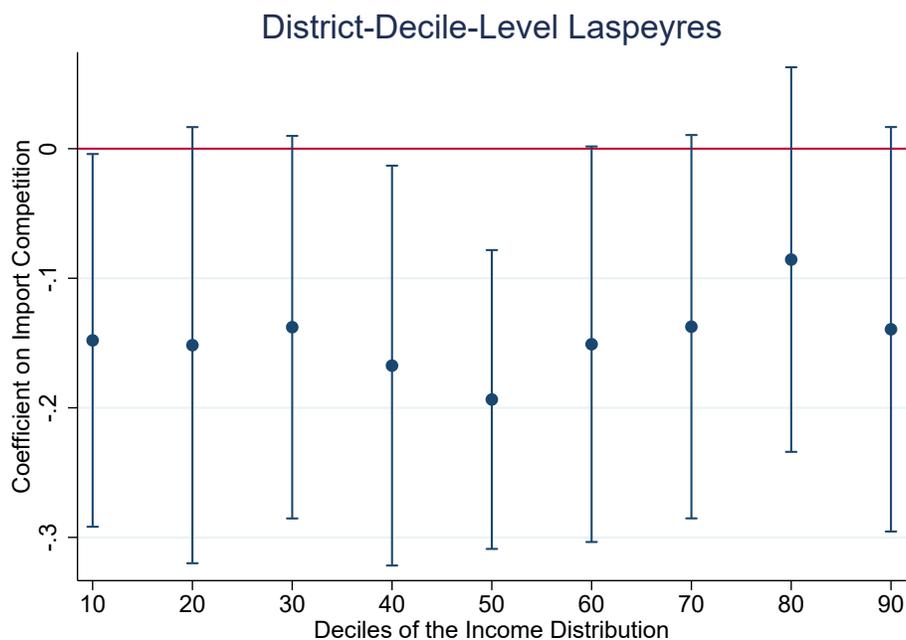
Notes: Figure replicates regression specification used in Figure 9, but with Laspeyres price index changes for food and fuels as the dependent variable (instead of nominal outlays per capita or welfare). Laspeyres price indices calculated using district-by-decile-specific budget shares. Positive point estimates indicate negative effects of import tariff cuts. See Section 6.3 for discussion.

Appendix Tables

Table A.1: Product Groupings

3 G groups	8 G groups	34 g goods	Disaggregated NSS survey items included in the g goods
Raw food products	Cereals	Cereals - rice	Rice; chira; khoi, lawa; muri; other rice products.
Raw food products	Cereals	Cereals - wheat	Wheat, atta, wheat/atta PDS; maida; suji, rawa; sewai (noodles); bread (bakery).
Raw food products	Cereals	Cereals - coarse	Jowar, jowar products; bajra, bajra products; maize, maize products; barley, barley products; small millets, small millets products; ragi, ragi products; other cereals.
Raw food products	Gram and pulses	Gram	Gram (full grain/whole); gram products.
Raw food products	Gram and pulses	Pulses - besan, moong	Besan; moong; soyabean; other pulse products.
Raw food products	Gram and pulses	Pulses - urd, masur	Urd; masur; arhar (tur); khasari; peas (dry); gram (split); other pulses.
Raw food products	Meat, fish and eggs	Meat	Goat meat, mutton; beef, buffalo meat; pork; poultry.
Raw food products	Meat, fish and eggs	Fish, prawn	Fish, prawn.
Raw food products	Meat, fish and eggs	Eggs	Eggs, egg products.
Raw food products	Fruits and vegetables	Vegetable - root vegetables	Potato; arum; radish; carrot; turnip; beet; sweet potato; onion; other root vegetables.
Raw food products	Fruits and vegetables	Vegetable - gourds	Pumpkin; gourd; bitter gourd; cucumber; parwal/patal; jhinga/torai; snake gourd; other gourds.
Raw food products	Fruits and vegetables	Vegetable - leafy vegetables	Cauliflower; cabbage; brinjal; lady's finger; french beans, barbati; tomato; palak/other leafy vegetables.
Raw food products	Fruits and vegetables	Vegetable - other vegetables	Peas (fresh); chilli (green); capsicum; plantain(green); jackfruit (green); other vegetables.
Raw food products	Fruits and vegetables	Premium Fruits	Apple; grapes; leechi; orange; mausani; pineapple; pears (naspati); mango; watermelon.
Raw food products	Fruits and vegetables	Other fresh fruits	Banana; jackfruit; singara; papaya; kharbooza; berries.
Raw food products	Fruits and vegetables	Dry fruits and nuts	Coconut (copra); groundnut; dates; cashewnut; walnut; raisin (kishmish, monacca, etc.).
Other food products	Dairy products and edible oils	Ghee	Ghee; butter.
Other food products	Dairy products and edible oils	Milk	Milk (liquid).
Other food products	Dairy products and edible oils	Other milk products	Milk (condensed/powder); curd; baby food.
Other food products	Dairy products and edible oils	Vanaspati, margarine	Vanaspati, margarine.
Other food products	Dairy products and edible oils	Edible oils	Ground nut oil; mustard oil; coconut oil; other edible oils.
Other food products	Sugar, salt, and spices	Sugar	Sugar; gur; sugar candy (misri); honey.
Other food products	Sugar, salt, and spices	Salt	Salt.
Other food products	Sugar, salt, and spices	Spices	Turmeric; black pepper; dry chillies; garlic; tamarind; ginger; curry powder; oil seeds.
Other food products	Refreshments and intoxicants	Beverages	Tea (leaf); coffee (cups); ice; coconut (green).
Other food products	Refreshments and intoxicants	Processed food	Cooked meals; pickles; sauce; jam, jelly.
Other food products	Refreshments and intoxicants	Pan	Pan (finished); supari; lime; katha.
Other food products	Refreshments and intoxicants	Tobacco	Bidi; cigarettes; leaf tobacco; snuff.
Other food products	Refreshments and intoxicants	Intoxicants	Country liquor; beer; foreign liquor or refined liquor.
Fuels	Fuels	Coke, coal, charcoal	Coke; coal; charcoal.
Fuels	Fuels	Kerosene	Kerosene.
Fuels	Fuels	Firewood and chips	Firewood and chips.
Fuels	Fuels	Electricity	Electricity.
Fuels	Fuels	Matches	Matches.

Notes: This table details the classification of disaggregated NSS items (column 4) into various levels of aggregation: the 34 g goods used in our baseline analysis (column 3); the 8 groups that form the basis of the alternative G groupings we explore in Section 6.1 (column 2); and the 3 G groups each g good is assigned to in our baseline analysis (column 1). Different disaggregated NSS items in column 4 are separated by a semicolon. NSS items exclude those dropped by Deaton (2003b). Some NSS items were not consistently classified over rounds. Specifically: (Concorded) Rice uses individual items from R43 {Rice; Paddy} and R55 {Rice; Rice PDS}. (Concorded) Wheat uses R43 {Wheat; Atta} and R55 {Wheat/Atta PDS; Wheat/Atta other sources}. (Concorded) Jowar and Jowar products uses R43 {Jowar; Jowar products} and R55 {Jowar, Jowar products}. (Concorded) Bajra and Bajra products uses R43 {Bajra; Bajra products} and R55 {Bajra, Bajra products}. (Concorded) Maize and Maize products uses R43 {Maize; Maize products} and R55 {Maize; Maize products}. (Concorded) Barley and Barley products uses R43 {Barley; Barley products} and R55 {Barley; Barley products} and R55 {Barley, Barley products}. (Concorded) Small millets and Small millets products uses R43 {Small millets; Small millets products} and R55 {Small millets, Small millets products}. (Concorded) Ragi and Ragi products uses R43 {Ragi; Ragi products} and R55 {Ragi, Ragi products}. (Concorded) Beef, buffalo meat uses R43 {Beef, Buffalo meat} and R55 {Beef/Buffalo meat}. (Concorded) Goat, mutton uses R43 {Goat, Mutton} and R55 {Goat/Mutton}. (Concorded) Fish, Prawn uses R43 {Fish fresh; Fish dry} and R55 {Fish, prawn}. (Concorded) Eggs, Egg products uses R43 {Eggs; Egg products} and R55 {Eggs}. Vegetable- Gourds includes R43 {Papaya (green)} and R55 {Other gourds}. Vegetable - leafy vegetables includes R43 {Palak; Other leafy vegetables} and R55 {Palak / other leafy vegetables}. (Concorded) Vanaspati, margarine uses R43 {Vanaspati; Margarine} and R55 {Vanaspati/margarine}. (Concorded) Edible oils includes R43 {Linseed oil, Palm oil, Refined oil, Gingelly (til) oil, Rapeseed oil} and R55 {Edible oils (other)}. (Concorded) Sugar uses R43 {Sugar (crystal)} and R55 {Sugar PDS; sugar (other sources)}. (Concorded) Salt uses R43 {Sea salt; other salt} and R55 {Salt}.

Table A.2: Changes in Recall Periods and Within-Group Budget Shares

7-day recall interaction	Dependent variable: Relative budget shares (34 g goods)			Dependent variable: Relative budget shares (136 disaggregated NSS goods)
	Coefficient	Standard error	t-stat	
7-day recall X Cereals - coarse	0.00281	0.00292	0.96	
7-day recall X Cereals - rice	0.00213	0.00139	1.54	
7-day recall X Cereals - wheat	0.00103	0.00149	0.69	
7-day recall X Coke, coal, charcoal	0.01053	0.00511	2.06	
7-day recall X Dry fruits and nuts	-0.00014	0.00066	-0.2	
7-day recall X Eggs	-0.00065	0.00054	-1.21	
7-day recall X Electricity	0.00029	0.00316	0.09	
7-day recall X Firewood and chips	0.00177	0.00212	0.84	
7-day recall X Fish, prawn	0.00042	0.00117	0.36	
7-day recall X Ghee	0.00146	0.00279	0.52	
7-day recall X Gram	0.00140	0.00070	1.99	
7-day recall X Intoxicants	-0.00194	0.00400	-0.48	
7-day recall X Kerosene	-0.00210	0.00256	-0.82	
7-day recall X Matches	0.00009	0.00076	0.12	
7-day recall X Meat	0.00060	0.00124	0.48	
7-day recall X Milk	0.00038	0.00179	0.21	
7-day recall X Other Fresh fruits	0.00037	0.00099	0.38	
7-day recall X Other milk products	-0.00081	0.00262	-0.31	
7-day recall X Pan	-0.00122	0.00120	-1.02	
7-day recall X Premium Fruits	0.00012	0.00065	0.18	
7-day recall X Pulses - Besan, Moong	0.00003	0.00059	0.05	
7-day recall X Pulses - Urd, Masur	-0.00012	0.00061	-0.2	
7-day recall X Tobacco	0.00289	0.00122	2.36	
7-day recall X Vanaspati, margarine	0.00164	0.00167	0.98	
7-day recall X Vegetable - gourds	0.00033	0.00050	0.67	
7-day recall X Vegetable - leafy vegetables	0.00053	0.00056	0.96	
7-day recall X Vegetable - other vegetables	0.00005	0.00036	0.15	
7-day recall X Vegetable - root vegetables	-0.00022	0.00063	-0.35	
7-day recall X beverages	-0.00055	0.00069	-0.8	
7-day recall X edible oils	0.00093	0.00150	0.62	
7-day recall X processed food	0.00183	0.00367	0.5	
7-day recall X salt	0.00060	0.00038	1.61	
7-day recall X spices	-0.00055	0.00090	-0.61	
7-day recall X sugar	0.00021	0.00084	0.25	
District X g good Fixed Effects	Yes			No
District X disaggregated item Fixed Effects	No			Yes
Household weights	Yes			Yes
F-stat schedule*goods=0	1.04			1.02
p-value schedule*goods=0	0.401			0.422
N	263663			384344

Notes: For questions regarding quantities and expenditures on food, pan, tobacco and intoxicants, the thin NSS round 54 (January-June 1998) randomized households between a 30-day and a 7-day recall period. Table tests whether reported relative budget shares (expenditure on good i divided by expenditures on all goods in good i's G group) change with the recall period used. Columns 1-3 report coefficient estimates, standard errors and t-statistics from regression of relative budget shares on a dummy for whether the household was surveyed with a 7 day-recall period interacted with each of the 34 i products (after including district-product fixed effects). A significant coefficient on the interaction indicates that the recall period affected relative consumption reports for that good. The bottom of the table reports the test of joint significance for all interactions. Column 4 repeats the exercise but for the 136 disaggregated goods rather than the 34 aggregated goods we use in our baseline. Given the large number of estimates, in this case we simply report the F-statistic and p-value for joint significance at the bottom of the table.

13

B Theory Appendix

B.1 Proof of Lemma 1

Denote $q_i(p^t, y_h^t)$ the Marshallian demand for good i, function of prices p^t at time t and household h income y_h^t. Denote $E_i(p^t, y) = p_i q_i(p, y)/y$ the Engel curve for good i as a function of income y for a given set of prices p_t, and denote $V(p^t, y_h^t)$ the indirect utility function. In Lemma 1, the key property that we exploit is that q_i, E_i and V are all homogeneous of degree zero in p, y.

The first step is to show that Engel curves shift uniformly by $+\log \lambda$ if we have log total outlays (income) on the horizontal axis. By definition, we have

$$E_i(p^0, \frac{y_h^1}{\lambda}) = \frac{p_i^0 q_i(p^0, y_h^1/\lambda)}{(y_h^1/\lambda)} = \frac{\lambda p_i^0 q_i(p^0, y^1/\lambda)}{y^1}$$

but given that demand is homogeneous of degree zero in p, y, we have $q_i(p^0, y^1/\lambda) = q_i(\lambda p^0, y^1)$ and thus we obtain:

$$E_i(p^0, \frac{y_h^1}{\lambda}) = \frac{\lambda p_i^0 q_i(\lambda p^0, y^1)}{y^1} = \frac{p_i^1 q_i(p^1, y_h^1)}{y_h^1} = E_i(p^1, y^1)$$

Using this property, we can then show that the horizontal shift of Engel curves reflects the changes in price index and welfare:

i) Define the price index relative to prices in period 0 implicitly as $P^1(p^0, p^1, y^1)$ such that: $V(p^1, y_h^1) = V(p^0, \frac{y_h^1}{P^1})$. With the homogeneous change in prices $p^1 = \lambda p^0$, it is immediate to verify that $P^1 = \lambda$ given that indirect utility is homogeneous of degree zero:

$$V(p^1, y_h^1) = V(\lambda p^0, y_h^1) = V(p^0, \frac{y_h^1}{\lambda})$$

Similarly, define the price index relative to prices in period 1 implicitly as $P^0(p^0, p^1, y^0)$ such that: $V(p^0, y_h^0) = V(p^1, \frac{y_h^0}{P^0})$. With the homogeneous change in prices $p^1 = \lambda p^0$, it is again immediate to verify that $P^0 = 1/\lambda$. Given that Engel curves shift by a factor λ, we obtain:

$$E_i(p^0, \frac{y_h^1}{P^1}) = E_i(p^0, \frac{y_h^1}{\lambda}) = E_i(p^1, y^1)$$

and

$$E_i(p^1, \frac{y_h^0}{P^0}) = E_i(p^1, \lambda y_h^0) = E_i(p^0, y^0)$$

hence the shift (in log) of the Engel curve from period 0 to period 1 corresponds to the price index change $\log P^1$, and the shift from period 1 to period 0 corresponds to the price index change $\log P^0$. This proves assertion i).

ii) Compensating variations CV_h are implicitly defined as $V(p^1, y^1 - CV_h) = V(p^0, y_h^0) = U_h^0$. With the homogeneous change in prices $p^1 = \lambda p^0$, we can verify that compensating variations CV_h are such that $y_h^1 - CV_h = \lambda y^0$:

$$V(p^1, y^1 - CV_h) = V(p^0, y^0) = V(p^1/\lambda, y^0) = V(p^1, \lambda y_h^0)$$

Given that Engel curves shift by a factor λ, we obtain:

$$E_i(p^1, y_h^1 - CV_h) = E_i(p^1, \lambda y_h^0) = E_i(p^0, y_h^0)$$

hence the initial observed expenditure share $p_i^0 q_{ih}^1/y_h^1$ of good i in period 0 corresponds to the counterfactual expenditure share of good i at new prices and total outlays $y_h^1 + EV_h$.

Equivalent variations EV_h are implicitly defined as $V(p^0, y^0 + EV_h) = V(p^1, y^1) = U_h^1$. For EV_h the proof proceeds the same way as for CV_h just by swapping periods 0 and 1 (and $1/\lambda$ instead of λ). With the homogeneous change in prices $p^1 = \lambda p^0$, we can verify that equivalent variations EV_h)

are such that $y_h^0 + EV_h = y^1/\lambda$:

$$V(p^0, y^0 + EV_h) = V(p^1, y^1) = V(\lambda p^0, y^1) = V(p^0, y_h^1/\lambda)$$

Again we obtain:

$$E_i(p^0, y_h^0 + EV_h) = E_i(p^0, y^1/\lambda) = E_i(p^1, y^1)$$

hence the new observed expenditure share $p_i^1 q_{ih}^1/y_h^1$ of good i corresponds to the counterfactual expenditure share of good i at former prices at $y_h^0 + EV_h$.

B.2 Proof of Lemma 2

Suppose that for a certain good i the shift of the Engel curve $E_i(p^t, y_h^t)$ (expenditure share x_{ih}^t/y_h^t plotted against total outlays y_h^t) reflects the price index change for any realization of price changes across periods and any y, i.e. $E_i(p^1, y) = E_i(p^0, y/P^1(y))$. We know already from Lemma 1 that this is true for any preferences if we impose the price changes to be uniform across goods: $p^1 = \lambda p^0$. For it to be true for all price changes, we show that it implies:

- **Step 1**: the expenditure share x_{ih}/y_h does not depend on prices, conditional on utility.
- **Step 2**: this expenditure share x_{ih}/y_h does not depend on utility either (i.e. the utility function has a Cobb-Douglas upper tier in i vs. non-i).

Step 1. Stating that the shifts in the Engel curve reflect the price index change means more formally that for any income level y_h^1, we have:

$$E_i(p^1, y_h^1) = E_i(p^0, y_h^1/P^1(y_h^1)) \tag{A.1}$$

where $P^1(y_h^1)$ is the price index change transforming income at period 1 prices to income in 0 prices. By definition, the price index change P^1 is such that $V(p^1, y_h^1) = V(p^0, y_h^1/P^1)$ where V denotes the indirect utility function. An equivalent characterization of the price index is:

$$\frac{y_h^1}{P^1(y_h^1)} = e(V(p^1, y_h^1), p^0) = e(U_h^1, p^0)$$

using the expenditure function e, denoting utility in period 1 by U_h^1. Looking at the share good i in total expenditures and imposing that Engel curves satisfy condition A.1, we can see that it no longer depends on prices p^1 once we condition on utility U_h^1:

$$\frac{x_{ih}}{y_h} = E_i(p^1, y_h^1) = E_i\left(p^0, \frac{y_h^1}{P^1(y_h^1)}\right) = E_i(p^0, e(U^1, p^0))$$

(note that the expenditure share at time 1 is independent of prices p^0 in another period).

Step 2. So from now on, denote by $w_i(U)$ the expenditure share of good i as a function of utility. Let us also drop the time superscripts for the sake of exposition. Here in step 2 we show that w_i must be constant for demand to be rational.

Suppose that relative prices remain unchanged among other goods $j \neq i$, but relative prices still vary between good i and the other goods. Using the composite commodity theorem (applied to non-i goods), the corresponding demand for i vs. non-i goods should correspond to a rational demand system in two goods. Hence we will do as if there is only one good $j \neq j$ aside from i. We will denote by p_j the price of this other good composite j.

A key (although trivial) implication of adding up properties is that the share of good j in expenditure is given by $1 - w_i(U)$ and only depends on utility. Denote by $e(p, U)$ the aggregate expenditure function. Shephard's Lemma implies:

$$\frac{\partial \log e(p, U)}{\partial \log p_i} = w_i(U) \qquad , \qquad \frac{\partial \log e(p, U)}{\partial \log p_j} = 1 - w_i(U)$$

Hence, conditional on utility U, the expenditure function is log-linear in log prices. Integrating, we get:

$$\begin{aligned} \log e(p, U) &= \log e_0(U) + w_i(U) \log p_i + (1 - w_i(U)) \log p_j \\ &= \log e_0(U) + w_i(U) \log(p_i/p_j) + \log p_j \end{aligned}$$

This must hold for any relative prices. Yet, the expenditure function must also increase with utility, conditional on any prices. Suppose by contradiction that there exist $U' > U$ such that $w_i(U') > w_i(U)$ (the same argument works with $w_i(U') > w_i(U)$). We can then find $\log(p_i/p_j)$ such that:

$$\log(p_i/p_j)) > \frac{\log e_0(U) - \log e_0(U')}{w_i(U') - w_i(U)}$$

which implies:

$$\log e_0(U) + w_i(U)\log(p_i/p_j) > \log e_0(U') + w_i(U')\log(p_i/p_j)$$

which contradicts the monotonicity of the expenditure function in U. Hence w_i is constant and we have a Cobb-Douglas expenditure function with a constant exponent, leading to a flat Engel curve for good i.

B.3 Proof of Proposition 1

Proposition 1 states that quasi-separability in group G is a necessary and sufficient condition for the shifts in within-G Engel curves to exactly reflect price index changes when relative prices do not change within group G. The proof that quasi-separability is a necessary condition relies on part i) of Lemma 3 that we prove in the next section.

Quasi-Separability as a Sufficient Condition. In brief, the intuition is that, thanks to the quasi-separability assumption, relative expenditures in i within group G only depend on the level of utility and within-group relative prices (we hold the latter constant). The first step is to show that quasi-separability implies a relationship as stated in condition i) of Lemma 3.

Quasi-separability in G implies that the expenditure function can be written:

$$e(p, U) = \tilde{e}(\tilde{P}_G(p_G, U), p_{NG}, U)$$

using Shephard's Lemma we obtain that compensated (Hicksian) demand for two goods $i \in G$ is:

$$h_i(p, U) = \frac{\partial e(p, U)}{\partial p_i} = \frac{\partial \tilde{e}(p, U)}{\partial P_G}\frac{\partial \tilde{P}_G(p_G, U)}{\partial p_i}$$

Taking the sum across goods in G, multiplying by prices and using the assumption that P_G is homogeneous of degree one: $\tilde{P}_G = \sum_i p_i \frac{\partial \tilde{P}_G(p_G, U)}{\partial p_i}$ (Euler's identity), we obtain:

$$\sum_{i \in G} p_i h_i(p, U) = \frac{\partial \tilde{e}(p, U)}{\partial P_G}\sum_i p_i\frac{\partial \tilde{P}_G(p_G, U)}{\partial p_i} = \frac{\partial \tilde{e}(p, U)}{\partial P_G}\tilde{P}_G$$

Looking at relative expenditures in i within group G, we get:

$$\frac{x_i}{x_G} = \frac{p_i h_i(p, U)}{\sum_{j \in G} p_j h_j(p, U)} = \frac{\partial \log \tilde{P}_G(p_G, U)}{\partial \log p_i} \equiv H_{iG}(p_G, U) \tag{A.2}$$

i.e. the expenditure share of good 1 within G depends only on utility u and the vector of prices p_G of goods that belong to group G. Note that compensated demand is homogeneous of degree zero in prices. Hence, using our assumption that relative prices remain constant: $p'_G = \lambda_G p_G$ across the goods of group G, we obtain:

$$H_{iG}(p'_G, U') = H_{iG}(p_G, U')$$

For a consumer at initial utility u, income y and price p, notice that:

$$E_{iG}(p, y) = H_{iG}(p_G, U)$$

Denoting indirect utility by $V(p, y)$, we obtain the key identity behind Proposition 1:

$$H_{iG}(p_G, V(p, y))) = E_{iG}(p, y) \tag{A.3}$$

which holds for any income y (and also any price p and subvector p_G).

Using this equality, we can now obtain each subpart i) and ii) of Proposition 1 on Engel curves:

i) For part i), define $P^1(p^0, p^1, y_h^1)$ the exact price index change at income y_h^1 for household h, implicitly defined such that $V(p^0, y^1/P^1) = V(p^1, y_h^1)$ where V is the indirect utility function. Using equality (A.3) and the assumption that relative prices remain constant within G: $p_G^1 = \lambda_G p_G^0$, We obtain:

$$
\begin{aligned}
E_{iG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) &= H_{iG}\left(p_G^0, V(p^0, y_h^1/P^1(p^1, p^0, y_h^1))\right) \\
&= H_{iG}(p_G^0, V(p^1, y^1)) \\
&= H_{iG}(p_G^1, V(p^1, y^1)) \\
&= E_{iG}(p^1, y_h^1)
\end{aligned}
$$

where we go from the second to third line by noticing that H_{iG} is homogeneous of degree zero in prices (and $p_G^1 = \lambda_G p_G^0$ for some scalar λ_G). By switching time superscripts 1 and 0, we prove a similar equality using the other price index $P^0(p^0, p^1, y_h^0)$:

$$
E_{iG}\left(p^1, y^0/P^0(p^0, p^1, y_h^0)\right) = E_{iG}(p^0, y_h^0)
$$

The shift from one to the other Engel curve is given by each price index (which may vary across income levels y_h), from period 0 to 1 and from 1 to 0.

ii) By definition, compensating variations CV_h satisfy:

$$
V(p^1, y_h^1 - CV_h) = V(p^0, y^0) = U_h^0
$$

where U_h^0 denotes the utility level of household h in period 0. With the definition of CV_h and the homogeneity of function H_{iG} described above, we obtain that CV_h satisfies:

$$
\begin{aligned}
E_{iG}(p^1, y_h^1 - CV_h) &= H_{iG}(p_G^1, V(p^1, y_h^1 - CV_h)) \\
&= H_{iG}(p_G^1, U_h^0) \\
&= H_{iG}(p_G^0, U_h^0) \\
&= x_{ih}^0/x_{hG}^1
\end{aligned}
$$

where the last term refers to the within-group G expenditure share of good i in period 0.

Similarly, by definition, equivalent variations EV satisfy:

$$
V(p^0, y_h^0 + EV_h) = V(p^1, y_h^1) = U_h^1
$$

where U_h^1 denotes to the period 1 utility level of household h.

With the definition of EV_h and the homogeneity of function H_{iG}, we obtain that EV_h satisfies:

$$
\begin{aligned}
E_{iG}(p^0, y_h^0 + EV_h) &= H_{iG}(p_G^0, V(p^0, y_h^0 + EV_h)) \\
&= H_{iG}(p_G^0, U_h^1) \\
&= H_{iG}(p_G^1, U_h^1) \\
&= x_{hi}^1/x_{Gh}^1
\end{aligned}
$$

where the last term refers to the within-group G expenditure share of good i in period 1.

Quasi-Separability as a Necessary Condition. The proof starts with the same argument as in Lemma 2: for the shifts in Engel curves to reflect the changes in price indices, we need within-G expenditure shares to depend only on utility and relative prices within group G. In a second step, we use part i) of Lemma 3 (proven in the following appendix section) to obtain that quasi-separability is required.

Stating that the shifts in relative Engel curve reflect the price index change means more formally that for any income level y_h^1:

$$
E_{iG}(p^1, y_h^1) = E_{iG}(p^0, y_h^1/P^1(y_h^1)) \tag{A.4}
$$

where $P^1(y_h^1)$ is the price index change transforming income at period 1 prices to income in 0 prices. By

definition of the price index, P^1 is such that $V(p^1, y_h^1) = V(p^0, y_h^1/P^1)$ where V denotes the indirect utility function. Or equivalently:

$$\frac{y_h^1}{P^1(y_h^1)} = e(V(p^1, y_h^1), p^0) = e(U_h^1, p^0)$$

using the expenditure function e, where we denote utility in period 1 by U_h^1. Looking at the share good i in expenditures within group G, and imposing that Engel curves satisfy condition A.4, we can see that it no longer depends on prices p^1 once we condition on utility U_h^1:

$$\frac{x_{ih}}{y_h} = E_i(p^1, y_h^1) = E_i\left(p^0, \frac{y_h^1}{P^1(y_h^1)}\right) = E_G(p^0, e(U^0, p^0))$$

Note that the expenditure share at time 1 is independent of prices p^0 in another period. Hence there exists a function H_{iG} of within-G relative prices and utility such that:

$$\frac{x_{ih}}{y_h} = H_{iG}(p_G, U_h)$$

This is condition i) of Lemma 3. As we prove below, this condition implies quasi-separability in G, as shown below in the proof of Lemma 3. Hence, quasi-separability in G is required if we want the shifts in relative Engel curves to reflect the changes in price indices.

B.4 Proof of Lemma 3

Gorman (1970) and Deaton and Muellbauer (1980) have already provided a proof of the equivalence between quasi-separability and condition ii), using the distance function. Here for convenience we provide a proof without referring to the distance function.

Blackorby, Primont and Russell (1978), theorem 3.4) show the equivalent between quasi-separability (which they refer to as separability in the cost function) and condition i). The proof that we provide here is more simple and relies on similar argument as Goldman and Uzawa (1964) about the separability of the utility function.

In the proof below, we drop the household subscripts and time superscripts to lighten the notation.

• **Quasi-separability implies i).** Actually we have already shown that quasi-separability implies i). In the proof of Proposition 1 above, we have shown in equation (A.2) that we have:

$$\frac{x_i}{x_G} = H_{iG}(p_G, U) = \frac{\partial \log \tilde{P}_G}{\partial \log p_i}$$

if the expenditure function can be written as $e(p, U) = \tilde{e}(\tilde{P}_G(p_G, U), p_{NG}, U)$ where $\tilde{P}_G(p_G, U)$ is homogeneous of degree one in the prices p_G of goods in G.

The most difficult part of the proof of Lemma 3 is to show that condition i) leads to quasi-separability:

• **i) implies quasi-separability.**

Let us assume (condition i) that the within-group expenditure share of each good $i \in G$ does not depend on the price of non-G goods:

$$\frac{p_i h_i(p, U)}{x_G(p, U)} = H_{iG}(p_G, U)$$

where $h_i(p, U)$ is the compensated demand and $x_G(p, U) = \sum_{j \in G} p_j h_j(p, U)$ is total expenditure in goods of groups G. As a first step, we would like to construct a scalar function $\tilde{P}_G(p_G, U)$ such that:

$$\frac{\partial \log \tilde{P}_G}{\partial p_i} = \frac{1}{p_i} H_{iG}(p_G, U) \tag{A.5}$$

for each i, and $\tilde{P}_G(p_{G0}, U) = 1$ for some reference set of prices p_{G0}. Thanks to the Frobenius Theorem used notably for the integrability theorem of Hurwicz and Uzawa (1971), we know that such problem admits a solution if and only if the derivatives $\frac{\partial(H_i/p_i)}{\partial p_j} = \frac{\partial(H_j/p_j)}{\partial p_i}$ are symmetric. We need to check that

this term is indeed symmetric for any two goods i and j in group G:

$$
\begin{aligned}
\frac{\partial(H_i/p_i)}{\partial p_j} &= \frac{\partial(h_i/x_G)}{\partial p_j} \\
&= \frac{1}{x_G}\frac{\partial h_i}{\partial p_j} - \frac{h_i}{x_G^2}\frac{\partial x_G}{\partial p_j} \\
&= \frac{1}{x_G}\frac{\partial h_i}{\partial p_j} - \frac{h_i}{x_G^2}\left[h_j + \sum_{g\in G}p_g\frac{\partial h_g}{\partial p_j}\right] \\
&= \frac{1}{x_G}\frac{\partial h_i}{\partial p_j} - \frac{h_i}{x_G^2}\sum_{g\in G}p_g\frac{\partial h_j}{\partial p_g} - \frac{h_i h_j}{x_G^2}
\end{aligned}
$$

where the last line is obtained by using the symmetry of the Slutsky matrix: $\frac{\partial h_i}{\partial p_j} = \frac{\partial h_j}{\partial p_i}$ for any i, j. Using the homogeneity of degree zero of the compensated demand w.r.t prices, we get: $\sum_{g\in G}p_g\frac{\partial h_i}{\partial p_g} = -\sum_{k\notin G}p_k\frac{\partial h_i}{\partial p_k}$ and thus:

$$
\begin{aligned}
\frac{\partial(H_i/p_i)}{\partial p_j} &= \frac{1}{x_G}\frac{\partial h_i}{\partial p_j} - \frac{h_i}{x_G^2}\sum_{g\in G}p_g\frac{\partial h_j}{\partial p_g} - \frac{h_i h_j}{x_G^2} \\
&= \frac{1}{x_G}\frac{\partial h_i}{\partial p_j} + \frac{h_i}{x_G^2}\sum_{k\notin G}p_k\frac{\partial h_j}{\partial p_k} - \frac{h_i h_j}{x_G^2}
\end{aligned}
$$

Given the symmetry of the Slutsky matrix, the first term $\frac{1}{x_G}\frac{\partial h_i}{\partial p_j}$ is symmetric in i and j, so is the third term. Using the assumption that $\frac{h_i}{h_j}$ does not depend on the price of non-G goods for any couple of goods $i, j \in G$ and $k \notin G$, we also obtain that the second term is symmetric in i and j: $h_i\frac{\partial h_j}{\partial p_k} = h_j\frac{\partial h_i}{\partial p_k}$ for any $k \notin G$. Hence:

$$
\frac{\partial(H_i/p_i)}{\partial p_j} = \frac{\partial(H_j/p_j)}{\partial p_i}
$$

and we can apply Frobenius theorem to find such a function $\tilde{P}_G$ satisfying equation A.5.

Note that $\sum_{i\in G}H_i(p_G, U) = 1$ for any price vector p_G and utility U, hence $\tilde{P}_G$ is homogeneous of degree one in p_G and can take any value in $(0, +\infty)$.

The second step of the proof is to show that the expenditure function depends on the price vector p_G only through the scalar function $\tilde{P}_G(p_G, U)$. To do so, we use the same idea as in Lemma 1 of Goldman and Uzawa (1964).[1] Using our constructed $\tilde{P}_G(p_G, U)$, notice that:

$$
\frac{\partial e}{\partial p_i} = \frac{\partial \tilde{P}_G}{\partial p_i} \cdot x_G(p, U) \tag{A.6}
$$

Since this equality is valid for any $i \in G$ and any value of $\tilde{P}_G$, it must be that the expenditure function e remains invariant as long as $\tilde{P}_G$ remains constant since the Jacobian of e w.r.t p_G is null whenever the Jacobian of $\tilde{P}_G$ is null. Hence e can be expressed as a function of $\tilde{P}_G$, utility U and other prices:

$$
e(p, U) = \tilde{e}(\tilde{P}_G(p_G, U), p_{NG}, U)
$$

This concludes the proof that i) implies quasi-separability.

- **ii) implies quasi-separability.** Suppose that utility satisfies:

$$
K\left(F_G(q_G, U), q_{NG}, U\right) = 1
$$

---

[1] Lemma 1 of Goldman and Uzawa (1964) states that if two multivariate functions f and g are such that $\frac{\partial f}{\partial x_i} = \lambda(x)\frac{\partial g}{\partial x_i}$ it must be that $f(x) = \Lambda(g(x))$ for some function Λ over connected sets of values taken by g.

Construct $\tilde{P}_G$ as follows:

$$\tilde{P}_G(p_G, u) = \min_{q_G} \left\{ \sum_{i \in G} p_i q_i \mid F_G(q_G, U) = 1 \right\}$$

which is homogeneous of degree 1 in p_G. Denote by $\tilde{e}$ the function of scalars P_G, U and price vectors p_{NG}:

$$\tilde{e}(P_G, p_{NG}, U) = \min_{Q_G, q_{NG}} \left\{ Q_G P_G + \sum_{i \notin G} p_i q_i \mid K(Q_G, q_{NG}, U) = 1 \right\}$$

The expenditure function then satisfies:

$$
\begin{aligned}
e(p, U) &= \min_{q_G, q_{NG}} \left\{ \sum_{i \in G} p_i q_i + \sum_{i \notin G} p_i q_i \mid K\left(F_G(q_G, U), q_{NG}, U \right) = 1 \right\} \\
&= \min_{q_G, Q_G, q_{NG}} \left\{ \sum_{i \in G} p_i q_i + \sum_{i \notin G} p_i q_i \mid F_G(q_G, U) = Q_G \; ; \; K(Q_G, q_{NG}, U) = 1 \right\} \\
&= \min_{q_G, Q_G, q_{NG}} \left\{ Q_G \sum_{i \in G} p_i q_i + \sum_{i \notin G} p_i q_i \mid F_G(q_G, U) = 1 \; ; \; K(Q_G, q_{NG}, U) = 1 \right\} \\
&= \min_{Q_G, q_{NG}} \left\{ Q_G \tilde{P}_G(p_G, U) + \sum_{i \notin G} p_i q_i \mid K(Q_G, q_{NG}, U) = 1 \right\} \\
&= \tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U \right)
\end{aligned}
$$

(going from the second to third lines uses the homogeneity of F_G) which proves that ii) implies quasi-separability.

• **Quasi-separability implies ii).** Now, assume that we have in hand two functions $\tilde{P}_G$ (homogeneous of degree 1) and $\tilde{e}$ that satisfies usual properties of expenditure functions. From these two functions, the goal is to:
    - implicitly construct utility that satisfies ii)
    - verify that $\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U \right)$ is the expenditure function associated with it.

First, using these two functions, let us define:

$$K(Q_G, q_{NG}, U) = \min_{P_G^*, p_{NG}^*} \left\{ \frac{Q_G P_G^* + \sum_{i \notin G} p_i^* q_i}{\tilde{e}(P_G^*, p_{NG}^*, U)} \right\} \tag{A.7}$$

and:

$$F_G(q_G, U) = \min_{p_G} \left\{ \frac{\sum_{i \in G} p_i^* q_i}{\tilde{P}_G(p_G^*, U)} \right\} \tag{A.8}$$

Those functions are similar to distance functions introduced by Gorman (1970). We can also check that both F_G and K are homogeneous of degree one in q_G. For instance, we have for F_G:

$$F_G(\lambda q_G, U) = \min_{p_G} \left\{ \frac{\sum_{i \in G} \lambda p_i^* q_i}{\tilde{P}_G(p_G^*, U)} \right\} = \lambda \min_{p_G^*} \left\{ \frac{\sum_{i \in G} p_i^* q_i}{\tilde{P}_G(p_G^*, U)} \right\} = \lambda F_G(q_G, U)$$

If $\tilde{e}$ and $\tilde{P}_G$ are decreasing in U, we can see that F_G and K are decreasing in U, hence the following has a unique solution:

$$K\left(F_G(q_G, U), q_{NG}, U \right) = 1 \tag{A.9}$$

Let us define utility implicitly as above. These implicitly defined preferences satisfy condition ii). The next step is to show that prices p^* that minimize the right-hand side of equations (A.7) and (A.8) also coincide with actual prices p. Then the final step is to show that the expenditure function coincides with $\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U \right)$.

Utility maximization subject to the budget constraint and subject to constraint (A.9) leads to the following first-order conditions in q_i:

$$\mu \frac{\partial K}{\partial Q_G} \frac{\partial F_G}{\partial q_i} = \lambda p_i \qquad if \ \ i \in G$$

$$\mu \frac{\partial K}{\partial q_j} = \lambda p_j \qquad if \ \ j \notin G$$

where p are observed prices and where μ and λ are the Lagrange multipliers associated with (A.9) and the budget constraints respectively. Using the envelop theorem, these partial derivatives are:

$$\frac{\partial K}{\partial Q_G} = \frac{P_G^*}{\tilde{e}\left(P_G^*, p_{NG}^*, U\right)} \quad ; \quad \frac{\partial K}{\partial q_j} = \frac{p_j^*}{\tilde{e}\left(P_G^*, p_{NG}^*, U\right)} \quad ; \quad \frac{\partial F_G}{\partial q_i} = \frac{p_i^*}{\tilde{P}_G\left(p_G^*, U\right)}$$

where P_G^* and p_i^* refer to counterfactual prices that minimize the right-hand side of equations (A.7) and (A.8) that define K and F_G. Note that these counterfactual prices may potentially differ from observed prices, but we will see now that relative prices are the same. Combining the FOC and envelop theorem, we obtain:

$$\mu \frac{P_G^*}{\tilde{e}\left(P_G^*, p_{NG}^*, U\right)} \frac{p_i^*}{\tilde{P}_G\left(p_G^*, U\right)} = \lambda p_i \qquad if \ \ i \in G$$

$$\mu \frac{p_j^*}{\tilde{e}\left(P_G^*, p_{NG}^*, U\right)} = \lambda p_j \qquad if \ \ j \notin G$$

But notice that if p_i^* for $i \in G$ minimizes the right-hand side of equation (A.8), then $\lambda_G p_i^*$ also minimizes (A.8) since $\tilde{P}_G$ is homogeneous of degree one. With $\lambda_G = \frac{\mu}{\lambda} \frac{P_G^*}{\tilde{e}\left(P_G^*, p_{NG}^*, U\right)} \frac{1}{\tilde{P}_G\left(p_G^*, U\right)}$, it implies that we can have: $p_i^* = p_i$ for $i \in G$. Also notice that if P_G^* and p_j^* for $j \notin G$ minimize the right-hand side of equation (A.7), then $\lambda_N P_G^*$ and $\lambda_N p_j^*$ also minimizes (A.8) for any $\lambda_N > 0$ since $\tilde{e}$ is homogeneous of degree one. With $\lambda_N = \frac{\mu}{\lambda \tilde{e}\left(P_G^*, p_{NG}^*, U\right)}$, we have $\lambda_N p_j^* = p_j$. Using the FOC for goods $j \notin G$, we obtain:

$$\frac{\mu}{\lambda} = \tilde{e}\left(\lambda_N P_G^*, p_{NG}, U\right)$$

In turn, the FOC for goods $i \in G$ yields:

$$\lambda_N P_G^* = \tilde{P}_G\left(p_G, U\right)$$

So we can also replace P_G^* by $\tilde{P}_G$.

Now that we have proven that observed prices are also solution of the minimization of (A.7) and (A.8), it is easy to show that $\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U\right)$ is equal to the expenditure function associated with utility defined in equation (A.9). Using equations (A.9), (A.7) and (A.8), and the equality between p^* and p (as well as P_G^* and P_G), we find:

$$
\begin{aligned}
\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U\right) &= F_G(q_G, U) P_G^* + \sum_{i \notin G} p_i^* q_i \\
&= F_G(q_G, U) P_G + \sum_{i \notin G} p_i q_i \\
&= \sum_{i \in G} p_i q_i + \sum_{i \notin G} p_i q_i
\end{aligned}
$$

where quantities are those maximizing utility subject to the budget constraint, therefore the expenditure function coincides with $\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U\right)$. Once we know that observe price minimize (A.7) and (A.8), it is also easy to verify that the expenditure shares implied by utility defined in A.9 also correspond to expenditure shares implied by the expenditure function $\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U\right)$. This shows that utility defined by (A.9), (A.7) and (A.8) leads to the same demand system as $\tilde{e}\left(\tilde{P}_G(p_G, U), p_{NG}, U\right)$, and proves that quasi-separability implies condition ii).

B.5 Proof of Proposition 2

As we have seen for the proof of Proposition 1, we have: $H_{iG}(p_G, V(p, y))) = E_{iG}(p, y)$ where $H_{iG}(p_G, U_h)$ denotes the within-G compensated expenditure share:

$$H_{iG}(p_G, U_h) = \frac{x_{hi}}{x_{hG}} = \frac{p_i h_i(p, U_h)}{\sum_{j \in G} p_j h_j(p, U_h)}$$

Denote utility in period 1 by $U_h^1 = V(p^1, y^1)$. We obtain:

$$
\begin{aligned}
E_{iG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) &= H_{iG}\left(p_G^0, V(p^0, y_h^1/P^1(p^1, p^0, y_h^1))\right) \\
&= H_{iG}(p_G^0, V(p^1, y^1)) \\
&= H_{iG}(p_G^1, V(p^1, y^1)) + [H_{iG}(p_G^0, U_h^1) - H_{iG}(p_G^1, U_h^1)] \\
&= E_{iG}\left(p^1, y_h^1\right) + [H_{iG}(p_G^0, U_h^1) - H_{iG}(p_G^1, U_h^1)] \\
&= E_{iG}\left(p^1, y_h^1\right) + \sum_{j \in G} \int_{\log p_j^1}^{\log p_j^0} \frac{\partial H_{iG}}{\partial \log p_j} d \log p_j
\end{aligned}
$$

where each step is similar to the those of the proof of Proposition 1, aside from the new term in the last three lines, re-expressed in the last line using the derivatives $\frac{\partial H_{iG}}{\partial \log p_j}$ evaluated along indifference curves at utility U_h^1.

Symmetric arguments can be used for P^0. This proves Proposition 2.

Is it possible for the econometrician to evaluate $\frac{\partial H_{iG}}{\partial \log p_j}$ without observing utility? To do so, one can use a Slutsky-type decomposition applied to within-G expenditure shares:

Lemma 4. *Compensated elasticities can be retrieved from a Slutsky decomposition:*

$$\frac{\partial H_{iG}}{\partial \log p_j} = \frac{\partial E_{iG}}{\partial \log p_j} + E_{jG} \frac{x_G}{y} \frac{\partial E_{iG}}{\partial \log y}$$

where $\frac{\partial E_{iG}}{\partial \log p_j}$ and $\frac{\partial E_{iG}}{\partial \log y}$ are the uncompensated elasticities which can be more directly estimated.

Proof Since $H_{iG}(p, U) = E_{iG}(p, e(p, U))$, and using $E_{jG} \frac{x_G}{y}$ the expenditure share of good j, we obtain:

$$\frac{\partial H_{iG}}{\partial \log p_j} = \frac{\partial E_{iG}}{\partial \log p_j} + \frac{\partial \log e}{\partial \log p_j} \frac{\partial E_{iG}}{\partial \log y} = \frac{\partial E_{iG}}{\partial \log p_j} + E_{jG} \frac{x_G}{y} \frac{\partial E_{iG}}{\partial \log y}$$

B.6 Proofs for Section 5.2

Orthogonality and First-Order Correction Terms for Relative Price Changes

First, note that the same result as in Proposition 2 holds in log. For instance, for P^1 we have:

$$
\begin{aligned}
\log E_{iG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) &= \log H_{iG}\left(p_G^0, V(p^0, y_h^1/P^1(p^1, p^0, y_h^1))\right) \\
&= \log H_{iG}(p_G^0, V(p^1, y^1)) \\
&= \log H_{iG}(p_G^1, V(p^1, y^1)) + [\log H_{iG}(p_G^0, U_h^1) - \log H_{iG}(p_G^1, U_h^1)] \\
&= \log E_{iG}\left(p^1, y_h^1\right) + [\log H_{iG}(p_G^0, U_h^1) - \log H_{iG}(p_G^1, U_h^1)] \\
&= \log E_{iG}\left(p^1, y_h^1\right) + \sum_{j \in G} \int_{\log p_j^1}^{\log p_j^0} \frac{\partial \log H_{iG}}{\partial \log p_j} d \log p_j \qquad \text{(A.10)}
\end{aligned}
$$

As a first-order approximation in the change in prices, we obtain:

$$\log E_{iG}(p^0, y^1/P^1) \approx \log E_{iG}(p^1, y_h^1) - \sum_{j \in G} \sigma_{ijh} \Delta \log p_j$$

$$= \log E_{iG}(p^1, y_h^1) - \sum_{j \in G} \sigma_{ijh}(\Delta \log p_j - \overline{\Delta \log p_G})$$

where $\sigma_{ijh} = \frac{\partial \log H_{iG}}{\partial \log p_j}$ is the compensated price elasticity of relative consumption of i with respect to price j, $\Delta \log p_j = \log p_j^1 - \log p_j^0$ is the change in the price of good j from the base period 0, and $\overline{\Delta \log p_G}$ is the average price change within G. Note that $\sum_{j \in G} \sigma_{ijht} = 0$ due to homogeneity of degree zero of H_{iG} in all G prices so we can rewrite $\sum_{j \in G} \sigma_{ijh} \Delta \log p_j$ as $\sum_{j \in G} \sigma_{ijh}(\Delta \log p_j - \overline{\Delta \log p_G})$. As a first-order approximation, we invert and obtain:

$$
\begin{aligned}
\frac{y_h^1}{P^1(y_h^1)} &= \log E_{iG}^{-1}\left(p^0, E_{iG}(p^1, y_h^1)\right) + (\beta_{ih}^0)^{-1} \sum_{j \in G} \sigma_{ijh}(\Delta \log p_j - \overline{\Delta \log p_G}) \\
&= \log E_{iG}^{-1}\left(p^0, \frac{x_{ih}^1}{x_{Gh}^1}\right) + (\beta_{ih}^0)^{-1} \sum_{j \in G} \sigma_{ijh}(\Delta \log p_j - \overline{\Delta \log p_G})
\end{aligned}
$$

where $\beta_{ihm}^0 = \frac{\partial \log E_{iG}}{\partial \log y_h}$ denotes the slope of the relative Engel curve, evaluated in period 0. Taking the average across goods, we obtain the expression in the text:

$$\log\left(\frac{y_h^1}{P^1}\right) \approx \frac{1}{G} \sum_{i \in G} \log E_{iG}^{-1}\left(p^0, \frac{x_{ih}^1}{x_{Gh}^1}\right) - \frac{1}{G} \sum_{i,j \in G} (\beta_{ih}^0)^{-1} \sigma_{ijh}(\Delta \log p_j - \overline{\Delta \log p_G})$$

Exact Correction Terms for Relative Price Changes

We can start again from Proposition 2 and expression (A.10), now imposing specific forms of demand. Suppose that demand is constant among goods within group G, i.e. such that:

$$H_{iG}(p_G, U) = \frac{A_i(U) p_i^{1-\sigma}}{\sum_{j \in G} A_j(U) p_j^{1-\sigma}}$$

If we have knowledge of the price elasticity σ and initial consumption shares, we can predict consumption shares for all goods i within G for any change in relative prices, holding utility constant:

$$
\begin{aligned}
H_{iG}(p_G', U) &= \frac{(p_i'/p_i)^{1-\sigma} A_i(U) p_i^{1-\sigma}}{\sum_{j \in G} (p_j'/p_j)^{1-\sigma} A_j(U) p_j^{1-\sigma}} \\
&= \frac{(p_i'/p_i)^{1-\sigma} H_{iG}(p_G, U)}{\sum_{j \in G} (p_j'/p_j)^{1-\sigma} H_{jG}(p_G, U)} \\
&= \frac{(p_i'/p_i)^{1-\sigma} (x_i/x_G)}{\sum_{j \in G} (p_j'/p_j)^{1-\sigma} (x_j/x_G)}
\end{aligned}
$$

For instance, to obtain P^1, applying the same logic as with Proposition 2 along with such a price adjustment yields:

$$
\begin{aligned}
E_{iG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) &= H_{iG}\left(p_G^0, V(p^0, y_h^1/P^1(p^1, p^0, y_h^1))\right) \\
&= H_{iG}(p_G^0, V(p^1, y^1)) \\
&= \frac{(p_i^0/p_i^1)^{1-\sigma} H_{iG}(p_G^1, V(p^1, y^1))}{\sum_{j \in G} (p_j^0/p_j^1)^{1-\sigma} H_{jG}(p_G^1, V(p^1, y^1))} \\
&= \frac{(p_i^0/p_i^1)^{1-\sigma} E_{iG}(p^1, y_h^1)}{\sum_{j \in G} (p_j'/p_j)^{1-\sigma} E_{jG}(p^1, y_h^1)}
\end{aligned}
$$

Another simple case is a special case of the EASI demand system (Lewbel and Pendakur, 2009). With

EASI, H_{iG} can be written as:

$$H_{iG}(p_G, U) = \frac{A_i(U) + B_i(p_G) + U D_i(p_G)}{\sum_{j \in G} A_j(U) + B_j(p_G) + U D_j(p_G)}$$

A specification that is linear in prices yields:

$$H_{iG}(p_G, U) = \frac{A_i(U) + \sum_{j \in G} B_{ij}(U) \log p_j}{\sum_{k \in G} A_k(U) + \sum_{k,j \in G} B_{kj}(U) \log p_j} = \frac{A_i(U) + \sum_{j \in G} B_{ij}(U) \log p_j}{\sum_{k \in G} A_k(U)}$$

since $\sum_{k \in G} B_{kJ}(U) = 0$ if preferences are required to be quasi-separable in group G. Price semi-elasticities are given by:

$$\xi_{iJ}(U) \equiv \frac{\partial H_{iG}}{\partial \log p_J} = \frac{B_{iJ}(U)}{\sum_k A_k(U)}$$

where $\sum_j \xi_{ij}(U) = 0$.

Conditional on initial expenditure shares and price semi-elasticities, we can again back out the change in expenditure shares for a given utility level:

$$\begin{aligned}
H_{iG}(p'_G, U) &= \frac{A_i(U) + \sum_j B_{ij}(U) \log p'_j}{\sum_k A_k(U)} \\
&= H_{iG}(p_G, U) + \frac{\sum_j B_{ij}(U)(\log p'_j - \log p_j)}{\sum_k A_k(U)} \\
&= H_{iG}(p_G, U) + \sum_j \xi_{ij}(U)(\log p'_j - \log p_j)
\end{aligned}$$

To obtain P^1, applying Proposition 2 yields:

$$\begin{aligned}
E_{iG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) &= E_{iG}\left(p^1, y_h^1\right) + [H_{iG}(p_G^0, U_h^1) - H_{iG}(p_G^1, U_h^1)] \\
&= E_{iG}\left(p^1, y_h^1\right) + \sum_j \xi_{ij}(\log p_j^0 - \log p_j^1)
\end{aligned}$$

As usual in the literature (e.g. Fajgelbaum and Khandelwal 2016), we could further specify that cross price elasticities are the same and equal to ξ/N, which leads to:

$$E_{iG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) = E_{iG}\left(p^1, y_h^1\right) - \xi \times (\Delta \log p_i - \overline{\Delta \log p_G}) \tag{A.11}$$

where $\overline{\Delta \log p_G}$ refers to the average log price change within group G.

Aggregation across Varieties of a Good

Suppose that group G of goods can be further partitioned into subgroups of goods (varieties): $G = g1 \cup g2 \cup g3...$ Denote by $E_{g,G}$ the expenditure share on subgroup g within group G. Under the assumptions of Proposition 1, we have for each variety: $E_{i,G}(p^1, y_h^1) = E_{i,G}(p^0, \frac{y_h^1}{P^1(y_h^1)})$, and $E_{i,G}(p^0, y_h^0) = E_{i,G}(p^{t1}, \frac{y_h^0}{P^0(y_h^0)})$. Taking the sum across varieties $i \in g$ of a subgroup g, we obtain:

$$E_{g,G}(p^1, y_h^1) = \sum_{i \in g} E_{i,G}(p^1, y_h^1) = \sum_{i \in g} E_{i,G}(p^0, \frac{y_h^1}{P^1(y_h^1)}) = E_{g,G}(p^0, \frac{y_h^1}{P^1(y_h^1)}) \tag{A.12}$$

and:

$$E_{g,G}(p^0, y_h^0) = \sum_{i \in g} E_{i,G}(p^0, y_h^0) = \sum_{i \in g} E_{i,G}(p^{t1}, \frac{y_h^0}{P^0(y_h^0)}) = E_{g,G}(p^{t1}, \frac{y_h^0}{P^0(y_h^0)}) \tag{A.13}$$

This proves the corollary.

Next, suppose that there exists a price index $P_g(p_g, U)$ summarizing prices for subgroup g, i.e. such that the expenditure function can be written: $e(p, U) = \tilde{e}(\tilde{P}(P_{g1}(p_{g1}, U), P_{g2}(p_{g2}, U), P_{g3}(p_{g3}, U), ...))$. In

24

this case, we can relax the assumption of Proposition 1: equations (A.12) and (A.13) above hold if we assume that relative price indices remain constant, i.e. $P_g^1(p_g, U) = \lambda_G P_g^0(p_g, U)$ instead of assuming that the relative prices of all varieties remain constant within group G. We can use these price indices in Proposition 2 instead of the prices of individual varieties.

To see this, first note that we can express within-G compensated expenditure shares across subgroups g as a function of prices indices $P_g(p_g, U)$ instead of the full vector of within-G prices p_G:

$$H_{g,G}(P_{g1}, P_{g2}, ..., U_h) = \frac{\sum_{i \in g} x_{hi}}{x_{hG}} = \frac{\partial \log \tilde{P}(P_{g1}, P_{g2}, ..., U)}{\partial \log P_g}$$

(see the proof of Lemma 3 and Proposition 1, equation A.2, for the derivation of compensated expenditure shares). Taking the sum across varieties $i \in g$, and using $H_{g,G}(P_G, V(p, y))) = \sum_{i \in g} H_{i,G}(p_G, V(p, y))) = \sum_{i \in g} E_{iG}(p, y) = E_{gG}(p, y)$, we obtain, as in Proposition 2:

$$
\begin{aligned}
E_{gG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) &= H_{g,G}\left(P_G^0, V(p^0, y_h^1/P^1(p^1, p^0, y_h^1))\right) \\
&= H_{g,G}(P_G^0, V(p^1, y^1)) \\
&= H_{g,G}(P_G^1, V(p^1, y^1)) + \sum_{g' \subset G} \int_{\log P_{g'}^1}^{\log P_{g'}^0} \frac{\partial H_{g,G}}{\partial \log P_{g'}} d \log P_{g'} \\
&= E_{gG}\left(p^1, y_h^1\right) + \sum_{g' \subset G} \int_{\log P_{g'}^1}^{\log P_{g'}^0} \frac{\partial H_{g,G}}{\partial \log P_{g'}} d \log P_{g'}
\end{aligned}
$$

where we use subgroup price indices P_g instead of individual prices p_G. By homogeneity of degree zero in subgroup price indices P_g, we obtain $H_{g,G}(P_G^0, V(p^1, y^1)) = H_{g,G}(P_G^1, V(p^1, y^1))$ if $P_g^1(p_g, U) = \lambda_G P_g^0(p_g, U)$, and thus $E_{gG}\left(p^0, y^1/P^1(p^0, p^1, y_h^1)\right) = E_{gG}\left(p^1, y_h^1\right)$.

Test of Quasi-Separability

Part i) of Lemma 3 shows that preferences are quasi-separable in G if and only if relative (compensated) expenditure shares x_i/x_G for any good $i \in G$ do not depend on the price of any good $j \notin G$ if we hold utility U constant:

$$\left. \frac{\partial \log(x_i/x_G)}{\partial \log p_j} \right|_U = 0$$

Instead, holding income constant (uncompensated), we obtain:

$$\left. \frac{\partial \log(x_i/x_G)}{\partial \log p_j} \right|_y = \frac{\partial \log(x_i/x_G)}{\partial \log U} \frac{\partial \log V}{\partial \log p_j} \tag{A.14}$$

where V denotes the indirect utility function. Using Roy's identity (in terms of elasticities):

$$\frac{\partial \log V}{\partial \log p_j} = -\frac{p_j q_j}{y} \frac{\partial \log V}{\partial \log y}$$

and substituting into equation A.14, we obtain:

$$\left. \frac{\partial \log(x_i/x_G)}{\partial \log p_j} \right|_y = -\frac{p_j q_j}{y} \frac{\partial \log(x_i/x_G)}{\partial \log U} \frac{\partial \log V}{\partial \log y} \tag{A.15}$$

where V is the indirect utility function. In turn, note that the elasticity of relative (uncompensated) expenditure share x_i/x_G w.r.t. income, holding prices constant, is:

$$\frac{\partial \log(x_i/x_G)}{\partial \log y} = \frac{\partial \log(x_i/x_G)}{\partial \log U} \frac{\partial \log V}{\partial \log y}$$

Substituting into equation A.15, we obtain our result which holds if and only if preferences are quasi-separable:

$$\frac{\partial \log(x_i/x_G)}{\partial \log p_j}\bigg|_y = -\frac{p_j q_j}{y}\frac{\partial \log(x_i/x_G)}{\partial \log y}$$

Note that it is possible to provide an alternative proof using Slutsky decomposition for good i and compare to the sum of other goods $i' \in G$.

Quasi-Separability and Misclassification

Suppose that preferences are not quasi-separable in G but QS in $G \cup \{0\}$ where good 0 denotes a good outside of G that we would have omitted in our estimation. Denote by $H_{iG}(p_G, p_0, U_h)$ the within-G compensated expenditure share. It not only depends on within-G prices but also on the price of good 0. Moreover, H_{iG} is now homogeneous of degree zero in (p_G, p_0) instead of p_G.

As a first-order approximation, we get that the bias is:

$$\begin{aligned}
\frac{1}{G}\sum_{i \in G}\log E_{iG}^{-1}\left(\frac{x_{ih}^1}{x_{Gh}^1}\right) &= \log\left(\frac{y_h^1}{P^1}\right) + \frac{1}{G}\sum_{i,j \in G}\left(\beta_{ih}^0\right)^{-1}\sigma_{ijh}(\Delta\log p_j - \overline{\Delta\log p_G}) \\
&\quad + (\Delta\log p_0 - \overline{\Delta\log p_G}) \times \frac{1}{G}\sum_{i \in G}\left(\beta_{ih}^0\right)^{-1}\sigma_{i0h}
\end{aligned}$$

(where the σ_{ijh}'s denote price elasticities). Hence we can see that there is no additional bias from omitting good 0 if either condition is satisfied:

- If the change in the price of good 0 is the same as the average change in pries among goods G:

$$\Delta\log p_0 = \overline{\Delta\log p_G}$$

- or if the price elasticity w.r.t good 0 is not correlated with the slope of the Engel curve among goods $i \in G$:

$$\frac{1}{G}\sum_{i \in G}\left(\beta_{ih}^0\right)^{-1}\sigma_{i0h} = 0$$

Let us now examine the case where a good j is wrongly classified as belonging to G while it should instead be treated as a non-G good.

Denote by $H_{jG}(p, U) = p_j h_j(p, U)/\sum_{j'} p_{j'}h_{j'}(p, U)$ the expenditure share in j within G (in terms of Hicksian demand), which now depends on the full vector of prices rather than just prices within G, but is still homogenenous of degree zero in prices. Again, as a first-order approximation leads to the following equality, now taking sums for log price changes across all goods k:

$$\log E_{jG}^{-1}\left(p^0, \frac{x_{jh}^1}{x_{Gh}^1}\right) \approx \log\left(\frac{y_h^1}{P^1}\right) + \left(\beta_{jh}^0\right)^{-1}\sum_k(\Delta\log p_k - \overline{\Delta\log p_G})\frac{\partial \log H_{jG}}{\partial \log p_k} \tag{A.16}$$

There is no bias if the relative price change across all goods k is not correlated with the difference in (compensated) price elasticities between j and G.

Heterogeneous Preferences

Here we examine the role of heterogeneity in preferences across demographic groups. Denote each group by an index k.

As a first simple case, assume that each group experience the same price index change for a given level of income (yet still heterogeneous across the income distribution). With a common change in price indices, the horizontal shift is the same across groups:

$$x_{iG,h,k}^1 = E_{iG,k}(p^1, y_h^1) = E_{iG,k}(p^0, \frac{y_h^1}{P^1(y_h^1)})$$

It is then easy to see that the average relative Engel curve across groups also shifts by $P^1(y_h^1)$, conditional on income y_h^1:

$$\overline{E}_{iG}(p^1, y_h^1) = \overline{E}_{iG}(p^0, \frac{y_h^1}{P^1(y_h^1)})$$

Hence, the average Engel curve across demographic groups we can still help us identify the price index change.

Now, suppose that $P_k^1(y_h^1)/P_{ref}^1(y_h^1) = 1 + \varepsilon_k^1(y_h^1)$. As a first-order approximation in ε, we obtain:

$$E_{iG,k}(p^1, y_h^1) = E_{iG,k}(p^0, \frac{y_h^1}{P^1(y_h^1)}) \approx E_{iG,k}(p^0, \frac{y_h^1}{P_{ref}^1(y_h^1)}) - \beta_{i,k}^1 \varepsilon_k^1$$

where $\beta_{i,k}^1(y_h^1)$ is the slope of the relative Engel curve for good i from period 1 for group k evaluated at income $y_h^1/P_{ref}^1(y_h^1)$ in log. Taking averages across groups, we obtain:

$$\overline{E}_{iG}(p^1, y_h^1) \approx \overline{E}_{iG}(p^0, \frac{y_h^1}{P_{ref}^1(y_h^1)}) - \frac{1}{K}\sum_k \beta_{i,k}^1 \varepsilon_k^1$$

If we use the average Engel curve, our estimated price index $\widetilde{P}^1$ is then such that:

$$\overline{E}_{iG}(p^0, \frac{y_h^1}{\widetilde{P}^1(y_h^1)}) \approx \overline{E}_{iG}(p^0, \frac{y_h^1}{P_{ref}^1(y_h^1)}) - \frac{1}{K}\sum_k \beta_{i,k}^1 \varepsilon_k^1$$

Inverting using the average relative Engel curve, this yields:

$$\log \widetilde{P}^1(y_h^1) \approx \log P_{ref}^1(y_h^1) + \frac{1}{K}\sum_k \beta_{i,k}^1 \varepsilon_k^1 / \bar{\beta}_i^{\,1}$$

where $\bar{\beta}_i^{\,1}$ denotes the average of the derivatives: $\bar{\beta}_i^{\,1} = \frac{1}{K}\sum_k \beta_{i,k}^1$ (and its inverse is equal to the derivative of the inverse of the average log Engel curve). If the price index is estimated by taking an average across goods, we obtain:

$$\log \widetilde{P}^1(y_h^1) \approx \log P_{ref}^1(y_h^1) + \frac{1}{K}\sum_k \beta_{i,k}^1 \varepsilon_k^1 / \overline{\beta_i^1}$$

This shows that we can interpret our naive estimator as a weighted estimator of heterogeneous price index changes, with weights proportional to $\sum_i \beta_{i,k}^1 / \overline{\beta_i^1}$.