The Representative Agent Bias in Cost of Living Indices

Sutirtha Bandyopadhyay[1] and Bharat Ramaswami[2]

Abstract

The aggregate cost of living index requires averaging across household indices. But what if the aggregate index was constructed for the 'representative' household as is usually done? The paper examines the resulting bias in the Tornqvist index, widely used for constructing superlative indices as well as for the Cobb–Douglas index, which has a similar functional form. We show that the representative agent index underestimates the aggregate index. We have shown that the bias depends on the heterogeneity in budget shares and the change in relative prices. Empirical application using Indian and US data, however, shows that the bias may be small and that the representative agent index is a good approximation to the aggregate index.

Keywords: Cost of Living Index, Aggregation, Heterogeneity, Representative Agent

[1]Visiting Assistant Professor, Indian Institute of Management Indore; Phone No: +917838757417; email: sutirtha.bandyopadhyay@gmail.com
[2]Professor, Ashoka University, Delhi, Phone No: +919811340848; email: isid.bharat@gmail.com

1. Introduction

This paper measures the representative agent bias in the construction of aggregate cost of living indices (COLIs). The paper considers the Tornqvist index, which is widely used for constructing superlative indices. The paper's results also apply to Cobb–Douglas indices, which are commonly used in theoretical and applied welfare analysis. While the results here are exposited for COLIs, they apply equally to Tornqvist indices of quantities and productivity.

Although the theory of COLIs is well developed for individual welfare, policy interest and practical questions have invariably been concerned with aggregate or group COLIs as a measure of changes in the welfare of that group. Given a Bergson–Samuelson social welfare function, Pollak (1981) showed that a group COLI could be defined in a fashion analogous to the individual COLI.[3]However, as Pollak points out, the premise that society has preferences that can be summarized by a social welfare function does not have universal acceptance.

A natural and more widely used definition is to consider the group COLI as an average of individual or household indices (Prais, 1959; Muellbauer, 1974; Nicholson, 1975; Pollak, 1980;Mackie and Schultze, 2002; Fisher and Griliches, 1995). The average can be unweighted (the so-called democratic index) or weighted, where the household indices are weighted according to that consumer's share of total expenditure (the so-called plutocratic index). The plutocratic index can also be rewritten as the ratio of the total expenditure required to enable each household to attain its reference period indifference curve at comparison prices to that required at reference prices. There has been some debate in the literature about whether the

[3]Another approach that also is based on a social welfare function is to let the social cost of living index be that uniform scaling of every individual's expenditure that keeps social welfare constant across a price change (Crossley and Pendakur, 2010).

aggregate index should be a democratic index or a plutocratic index. A democratic index, it is argued, is more representative because it weights poor and rich consumers equally.

Previous research has highlighted several knotty issues in the aggregation of household COLIs.[4] Households are heterogeneous with respect to spending patterns. Plutocratic indices, which are the ones typically reported by statistical agencies, are more representative of the consumption patterns of the higher income groups. Research has called for remedies either in terms of indices for sub-populations or a democratic aggregate index. Households may also be heterogeneous with respect to prices. However, if the statistical system is such that the price data is collected at the retail level, then it is those prices (which are in effect averaged across households) that are used rather than household-specific prices. The resulting index does not correspond to the theoretical notion of the aggregate COLI as the average of household COLIs.

In this paper, we consider the aggregation problem posed by heterogeneity in household budget shares. We abstract from the heterogeneity in prices; that is, we assume all households face the same prices. Even so, the aggregate COLI is an average of the household COLIs. We ask what the bias would be if the COLI was, instead, computed for a representative agent. The representative agent COLI would be the COLI that corresponds to average spending patterns. From previous work, we know that unless the expenditure function is of the polar Gorman form, a representative agent analysis is an invalid representation of the aggregate (Mackie and Schultze, 2002; Deaton and Muellbauer, 1980). The contribution here is to assess the direction and magnitude of bias for the important cases of the Tornqvist and Cobb–Douglas indices.

[4]For an overview of these issues, see Mackie and Schultze (2002).

The motivation for the question comes from the following. Despite the considerable theoretical appeal of COLIs, it is the Laspeyres price index that is usually reported by statistical agencies. The household Laspeyres index is linear in budget shares. Therefore, the aggregate Laspeyres index is also the Laspeyres index evaluated at the economy-wide average budget share. Hence, it is not necessary to compute the average of household-specific Laspeyres indices. The United States Bureau of Labor Statistics (BLS) carries over the practice of using the economy-wide budget shares in its construction of the Tornqvist index. However, this will certainly lead to bias because of the representative agent theorem. The Tornqvist index is non-linear in budget shares and it is, therefore, easy to see that the average of household indices will not be the index evaluated at the average. The extent of bias is an open question and that is addressed in this paper. The non-linearity of the Tornqvist index also raises a broader question: does the extent of heterogeneity in budget shares affect the Tornqvist COLI? The same question is valid for the Cobb–Douglas price index that resembles the Tornqvist index in functional form.

We are not aware of any other country reporting a superlative COLI. But if they plan to go in that direction, then they too must make a choice between using a representative analysis or computing the average of household indices. The temptation to use average budget shares and compute a representative agent COLI is understandable. Household-level COLIs require household-level budget shares as well as household-level price changes. Collecting data on the latter is a formidable task and agencies therefore rely on retail price data (Mackie and Schultze, 2002). The immense difficulty of accounting for price heterogeneity might lead statistical agencies also to ignore the other dimension of heterogeneity: in budget shares.

Our interest in the Tornqvist index comes from the fact that it is a superlative index (i.e. generated from an expenditure function of flexible form) that is derived from a non-homothetic translog expenditure function. The consistency with non-homothetic preferences endows the Tornqvist index with wide applicability. The Cobb–Douglas form is similar to the Tornqvist index and so the aggregation bias analysis easily extends to it.

The evaluation of an aggregate cost of living is essential to welfare analysis in many contexts and our motivating question can be posed in those situations as well. Consider the welfare effects of trade liberalization. A natural metric to measure the change in welfare is to look at the compensating variation (due to the change in trade policy) as a proportion of initial expenditure (e.g. Porto, 2006). But this is the same as the COLI (between the pre-liberalization and post-liberalization prices) minus one. Here again, the correct measure for aggregate welfare change would be an average of individual welfare changes. But what if average individual characteristics are used to evaluate the welfare change? What would be the bias? If the individual utility/welfare functions are Cobb–Douglas, then we can characterize the bias from the results stated in this paper.

A preview of our findings is as follows. The Tornqvist index and the Cobb–Douglas index are convex in budget shares. As a result, the greater is the heterogeneity in budget shares, the higher is the value of the COLI. It also means, for a given change in prices, that the aggregate COLI is greater than the representative agent COLI. The paper shows that the bias depends on the heterogeneity in budget shares as well as the extent of change in relative prices. We evaluate the magnitude of the bias corresponding to the Cobb–Douglas price index using expenditure data from the United States as well as India under different scenarios of price changes. The bias in the Tornqvist index is computed from panel data for the United States.[5] A similar computation for

[5]The panel data is constructed by Blundell, Pistaferri and Preston (2008) based on longitudinal and cross-sectional data.

India is not possible because of the unavailability of panel data. Indeed, as panel consumption data is not available for most countries, we also derive upper bounds for the bias in the Tornqvist index that can be estimated from cross-sectional data alone.

The paper finds the magnitude of the bias to be very small for both the Cobb–Douglas index and the Tornqvist index, which suggests that in practice the bias can be ignored. The upper bound to the bias in the Tornqvist index also turns out to be quite small and similar in magnitude to the bias estimated for the Cobb–Douglas index.

2. Relation to Literature

The officially reported COLIs by statistical agencies are price indices which usually measure the change in the cost of a fixed basket of goods and services as prices change. These fixed basket indices are limited measures of the true cost of living, as they fail to capture the substitution effect due to relative price changes.

Superlative indices are superior as they capture the substitution effect which occurs due to the change in relative prices (Manser and Mcdonald, 1988; Abraham *et al*., 1998; Boskin *et al*., 1998). Superlative indices provide a close approximation to a COLI using only the observable price and quantity data; that is, it would not be necessary to econometrically estimate the elasticities of substitution of all of the items with each other. The most widely known index number formulas that belong to the superlative class identified by Diewert are the Fisher Ideal index and the Tornqvist index. The Fisher index and the Tornqvist index are found to be close approximations of each other (Diewert, 1978; Dumagan, 2002). Apart from being a superlative index, the Tornqvist index has another interesting feature. It originates from an expenditure function that corresponds to non-homothetic preference (Diewert, 1976). Besides measuring the change in the cost of living, the Tornqvist functional form is widely used to measure the change in input, output and productivity (Caves *et al*., 1982).

Previous research has also clarified the notion of an aggregate COLI. The analogy from individual COLIs would suggest that it should be defined in a similar manner – as the ratio of expenditure required, at current prices, to meet a reference level of social welfare relative to the expenditure required, at reference period prices (Pollak,1981). However, the difficulty of specifying social welfare makes this approach impractical. Much of the literature therefore considers the aggregate index as the average of household indices (Prais, 1959; Muellbauer, 1974; Nicholson, 1975; Pollak, 1980; Mackie and Schultze, 2002; Fisher and Griliches, 1995).

The interpretability of such an average has, however, been questioned. The review of price indices by the panel of the National Academy of Science pointed out the difficulty: A single price index must somehow represent the average experience of a very heterogeneous population, whose members buy different goods, of different qualities, at different prices, in different kinds of outlets and who exhibit different substitution behavior when relative prices change' (Mackie and Schultze, 2002). Aggregation by way of an unweighted average –that is, a democratic index – reduces the bias that exists in a plutocratic index towards the consumption patterns of the better-off. However, a democratic index requires computation of household COLIs for a representative sample of households. Statistical agencies are not set up to do this, because while budget shares are drawn from household samples, they are combined with retail price data and therefore miss out on household heterogeneity in prices paid.

Household heterogeneity in budget shares has been emphasized by a number of papers that have examined the variation in household-specific COLIs and household-specific inflation rates(Michael, 1979; Hagemann, 1982; Idson and Miller, 1994; Crawford,1994; Crawford and Smith, 2002; Del Río and Ruiz-Castillo, 2002; Cage *et al*., 2002; Garner*et al*., 2003; Kokoski,1987; Garner *et al*., 1996; Livada,1990).Most of these papers track the difference between nominal and real expenditure inequality using these household-specific indices. Some of the papers also construct price indices for different sub-groups of the population like the elderly (Hobijn and Lagakos, 2003; Stewart, 2008) and for different demographic and income groups (Lyssiotou and Pashardes, 2004; Kokoski, 1987). All the papers mentioned assume varying spending patterns across households as the only source of heterogeneity. Prices faced by each household are assumed to be the same. Kaplan and Schulhofer-Wohl (2017) explore price heterogeneity for US households using scanner data (for another application of scanner data to construct a price index, see Prud'homme*et al*., 2005). The variation in household-specific COLIs constructed by these authors comes from heterogeneity in spending patterns as well as price heterogeneity.

Relative to this literature, our paper poses a different problem in aggregation. Like much of the heterogeneity literature, we assume all households face the same prices and are heterogeneous only in budget shares, which then is the only source of variation in the household COLI. The aggregate index is the average (plutocratic or democratic) of these household COLIs. However, if statistical agencies followed the practice of using average budget shares, they would arrive at the COLI of the average or representative agent. How well does this approximate the aggregate COLI?

As mentioned earlier, our analysis considers the Tornqvist index. The US BLS calculates the Tornqvist index regularly as an alternative consumer price index (CPI) in order to track the substitution bias in the fixed basket CPI. However, the calculation computes country- and region-specific Tornqvist indices that are representative in nature and hence suffer from the bias generated by individual heterogeneity.

The bias that occurs due to individual heterogeneity has deeper implications in applied welfare analysis. The application is not only limited to specific indices like the Tornqvist, which is used by statistical agencies and index number researchers. The functional form of the COLI derived from the Cobb–Douglas utility function is exactly similar to the Tornqvist and hence we can characterize the representative agent bias in a similar way.

In classical trade models (like the Heckscher–Ohlin model), we assume all consumers are homogeneous within a country and represent the welfare of the representative consumer by a Cobb–Douglas utility function. The equilibrium prices of commodities are determined within the model. The equilibrium prices differ before and after trade. Therefore, the cost of living differs between free trade and autarky. If we measure the change in the cost of living for a representative Cobb–Douglas consumer, it suffers from bias for not considering individual heterogeneity.

3. Does Heterogeneity Matter?

Consider a population of *N* households. We measure the change in the cost of living for each household by a Tornqvist index defined over *M* commodities.[6]For the 'j' th household, let $s_i^{1,j}$and $s_i^{0,j}$ be the budget shares for the *i*th commodity at period 1 and period 0, respectively. Define the average budget share as

$$s_i^j = \left(\frac{1}{2}\right)\left(s_i^{1,j} + s_i^{0,j}\right) \forall i = 1,2, \dots, M \ \& \ j = 1,2, \dots, N$$

Then the Tornqvist index for the *j*th household is

$$T^j\left(s_1^j, s_2^j, \dots, s_M^j\right) = \left(\frac{P_1^1}{P_1^0}\right)^{s_1^j}\left(\frac{P_2^1}{P_2^0}\right)^{s_2^j}\left(\frac{P_3^1}{P_3^0}\right)^{s_3^j} \dots \left(\frac{P_M^1}{P_M^0}\right)^{s_M^j}$$

All households face the same change in prices for all commodities, but budget shares vary across households.

Without loss of generality, assume the *M*th commodity to be the numeraire commodity. We denote $\lambda_i$to be the ratio of the relative price of commodity *i* in period 1 to its relative price in period 0; that is

$$\frac{P_i^1}{P_M^1} \Big/ \frac{P_i^0}{P_M^0} = \lambda_i \ \forall i = 1,2, \dots, M$$

---

[6]The Tornqvist index is generated from a flexible and non-homothetic translog expenditure function (Diewert, 1976). The expenditure function for the *j*th household is of the following form: $lnC^j(u,P) = a_0^j + \sum_{i=1}^M a_i^j \, lnP_i + \left(\frac{1}{2}\right)\sum_{i=1}^M \sum_{k=1}^M a_{ik}^j \, lnP_i \, lnP_k + b_0^j lnu^j + \sum_{i=1}^M b_i^j \, lnP_i lnu^j + \left(\frac{1}{2}\right) b_{00}(lnu^j)^2$. The parameters satisfy the following restrictions:$a_{ik}^j = a_{ki}^j \ \forall i = 1,2, \dots, M \ , k = 1,2, \dots, M \ and \ j = 1,2, .., N$;$\sum_{i=1}^M a_i^j = 1$; $\sum_{i=1}^M b_i^j = 0$;$\sum_{k=1}^M a_{ik}^j = 0 \forall i = 1,2, \dots, M \ and \ j = 1,2, .., N$.

Note that $(\lambda_i - 1)$ becomes the percentage change in the relative price of commodity *i*. Without loss of generality, we normalize the price ratio of commodity *M* between period 1 and period 0 to be one; that is, $\frac{P_M^1}{P_M^0} = 1$. Then using the fact that commodity budget shares sum to one, the Tornqvist index can be expressed as

$$T^j\left(s_1^j, s_2^j, \dots, s_M^j\right) = \prod_{i=1}^{M-1} \lambda_i^{s_i^j}$$

The aggregate index for this population is the average of the household Tornqvist indices. The average can be unweighted (democratic) or weighted (plutocratic). In either form, the aggregate index can be expressed as the expected value of the index over the households in the population. The democratic and plutocratic indices will, however, differ in the probability weights. This can be shown as follows.

Let $\boldsymbol{s}$ denote a particular allocation of budget shares, $(s_1, s_2, \ldots, s_M)$. Let $\mathrm{B} = \{\boldsymbol{s}\colon \sum_{m=1}^{M} s_m = 1\}$ denote the set of all possible allocations of budget shares. If $\boldsymbol{s}^j$ denotes the budget share allocation of the *j*th household, define the indicator function:

$$V^j(\boldsymbol{s}) = 1 \ if \ \boldsymbol{s}^j = \boldsymbol{s} \ \forall \ j = 1,2, \dots, N \ and \ \forall \ \boldsymbol{s} \in \mathrm{B}$$

$$V^j(\boldsymbol{s}) = 0 \ otherwise$$

The proportion of households that have the budget share allocation $\boldsymbol{s}$ is then given by

$$h(\boldsymbol{s}) = \left(\frac{1}{N}\right) \sum_{j=1}^{N} V^j(\boldsymbol{s})$$

Hence, the democratic aggregate COLI is defined by

$$A_d \equiv \sum_{s \in B} T(\boldsymbol{s})\, h(\boldsymbol{s}) = E(T(\boldsymbol{s}))$$

The corresponding representative agent index is

$$R_d = T(E(\boldsymbol{s}))$$

where $E(\boldsymbol{s})$ is the vector of average budget shares computed by using the density $h(\boldsymbol{s})$.

For the plutocratic group COLI, we define density as

$$k(\boldsymbol{s}) = \sum_{j=1}^{N} \left(\frac{C^j}{\sum_{j=1}^{N} C^j}\right) V^j(\boldsymbol{s}) \;\; \forall\, \boldsymbol{s} \in B$$

where C^j is the total expenditure made by the jth household.The plutocratic group cost of living index is defined as

$$A_p = \sum_{s \in B} T(\boldsymbol{s}) k(\boldsymbol{s}) = E(T(\boldsymbol{s}))$$

The corresponding representative agent index is

$$R_p = T(E(\boldsymbol{s}))$$

where $E(\boldsymbol{s})$ is the vector of average budget shares computed by using the density $k(\boldsymbol{s})$.

Whether democratic or plutocratic, the difference between the aggregate Tornqvist index and that of the representative agent is $E(T(\boldsymbol{s})) - T(E(\boldsymbol{s}))$ and therefore depends on the curvature of the Tornqvist index.

Proposition 1:$T(\boldsymbol{s})$ is convex in$\boldsymbol{s}$.

A proof is offered in the appendix. By Jensen's inequality, it follows that

Proposition 2: $E[T(\boldsymbol{s})] \geq \mathrm{T}[\mathrm{E(s)}]$

This result shows that representative agent approximation will underestimate the aggregate COLI. The convexity of the Tornqvist index has a further implication. An increase in heterogeneity in budget shares, in the sense of a Rothschild–Stiglitz mean-preserving spread (Rothschild and Stiglitz, 1970), increases the aggregate COLI.

The functional form of the COLI derived from the Cobb–Douglas utility function is exactly the same as the Tornqvist index (except for the fact that the budget share used is the same for the base and current periods). Therefore, propositions 1 and 2 also apply to the Cobb–Douglas price index.

We now turn to the second issue of determining the magnitude of bias because of the representative agent approximation.

Proposition 3: The representative agent bias can be approximated by the following:

(1) $g = \frac{E[T(\boldsymbol{s})] - T[\mathrm{E}(\boldsymbol{s})]}{T[\mathrm{E}(\boldsymbol{s})],} \approx \left(\frac{1}{2!}\right) var\left[\sum_{i=1}^{M-1} s_i \ln \lambda_i\right]$

where var stands for variance.

For a proof of this result, see the appendix to this paper. The expression in (1) is clearly non-negative. The representative agent bias is zero if there is no heterogeneity in the budget share. It is also zero when there is no change in relative prices, for then $\lambda_i = 1$.[7]Computing the bias in the

[7] Recall that the percentage rate of change in the relative price of the ith commodity is given by $(\lambda_i - 1)$.

Tornqvist index requires panel data at the household level to get information about the base and current period budget shares.

The counterpart of equation (1) for the Cobb–Douglas price index is

(2) $g \approx \left(\frac{1}{2!}\right) var\left[\sum_{i=1}^{M-1} \alpha_i \ln \lambda_i\right]$

where α_i is the base period (period 0) budget share. The bias in (2) can be estimated from cross-sectional data alone.

4. Representative Agent Bias in the Cobb–Douglas Index

We begin by presenting the bias estimates for the Cobb–Douglas index (i.e. equation (2)). For this purpose, we use cross-sectional data from India and the United States.

India

The nationally representative consumer expenditure survey of 2004-05 is used which samples about 120,000 households across rural and urban India. Following Almås and Kjelsrud (2017), we classify all expenditure into 11 categories. Tables 1 and 2 list these categories and also display across the urban and rural sectors, the mean budget shares as well as measures of dispersion. Notice that the coefficient of variation is more than 100% or close to 100% for few of the commodities. Such heterogeneity is not peculiar to the Indian data set.[8]

[8]In his study on the United States, Michael (1979) explains that the greater is the absolute variation in COLIs across households, the larger is the variance across households in the share of each item in the consumption bundle. Hobijn and Lagakos (2003) construct an experimental price index for the elderly in the US and find that between 1984 and 2001, the increase in the price index for the elderly was on average 0.38% higher than it was under the officially reported CPI by the BLS, with medical care accounting for much of the difference (share of medical expenditure turned out to be more than double for the elderly as compared to the overall population). Similarly, Garner *et al*.

Three scenarios of relative price changes (represented in Tables 3 and 4) are considered. In scenario 1, we consider the observed change in relative prices (relative to miscellaneous non-food, which is considered as a numeraire good) for all categories between 2011–12 and 2004–05.[9] In scenario 2, we suppose the percentage price changes are highest for the commodities consumed largely by the poor.[10] The prices of these categories are assumed to increase at a rate of 80%. Prices of all other categories are assumed to increase at a rate of 20% (including miscellaneous non-food). Scenario 3 is the exact opposite of scenario 2, where the prices of the most frequently consumed food groups by the poor increase by 20% and the prices of other categories increase by 80%. All these three scenarios can be compared with the benchmark scenario when there is no change in relative prices.

The bias is obviously zero for the benchmark scenario where the prices of all categories increase at the same rate. However, the representative agent bias also turns out to be very small for all the other three scenarios. This turns out to be true for the rural as well as the urban sample (shown in Tables 3 and 4). For the rural sample, the bias turns out to be 0.06%, 0.07% and 0.07% in scenarios 1, 2 and 3, respectively. The bias in the urban sample is 0.05% in scenario 1. The representative agent bias for the urban sample turns out to be 0.06% in scenario 2 as well as in scenario 3.

United States

(1996) construct an experimental price index for the poor, as the spending pattern for the poor is quite different from that for the rich. Crawford (1994) shows that budget share varies widely between the richest 10% and poorest 10% households for the UK and that causes the COLI to be different for these two groups. Del Rio and Ruiz-Castillo (2002) show high variation in budget shares for Spain and relate this variation to demographic and other characteristics of households.

[9]For the non-food categories, the observed changes are derived from changes in the corresponding components of the CPI. This cannot be done for the food categories, as the CPI does not provide it at the level of disaggregation considered in this paper. For this reason, the change in prices of food categories is derived from the changes in average unit value computed from the household expenditure survey.

[10]These are the food categories of ‘cereals and cereal substitutes’ and ‘pulse and pulse products’.

We use the US consumer expenditure survey (CEX) and divide consumption expenditure into four groups: ‘food consumed at home’, ‘food consumed away from home’, ‘other non-durable expenditure’ and ‘expenditure on all other goods’ (including durables, health and education). Table 5 shows the mean, standard deviation and compensating variation of the budget share of these categories for the year 1982. Combining the variance in budget share (in 1982) and change in relative prices (between 1992 and 1982), we compute the representative agent bias. The prices are yearly indices derived from the BLS.[11] The indices are US city averages, not seasonally adjusted, and are expressed in 1982–84 dollars.

Table 6 shows the change in relative prices and representative agent bias for three scenarios that are parallel to those considered in the Indian case. Scenario 1 represents the observed price change. In scenario 2, prices of the food categories (i.e. ‘food consumed in home’ and ‘food consumed away from home’) increase at a rate of 80% and prices of the other commodities increase at a rate of 20%. In the third case, food prices rise by 20% while the other prices rise by 80%.

Similar to the Indian case, the bias turns out to be small. We also compute the bias for a more disaggregated classification of commodities. Here we follow Johnson *et al*. (2006), who classify all non-durable and semi-durable expenditure into ten groups/categories.[12] However, BLS price data is not available for these categories. Hence, the bias is computed only for hypothetical price scenarios. In scenario 1, the prices of ‘food at home’, ‘food away from home’ and ‘utilities’ rise by 80%, while other prices increase by 20%. In scenario 2 the reverse happens – other prices increase by 80%, while those of the food categories and utilities increase by 20%.

[11] We source this information from Blundell *et al*. (2008).

[12] These are ‘food at home’, ‘food away from home’, ‘alcoholic beverages’, ‘utilities’, ‘personal care and miscellaneous goods’, ‘gas, motor fuel and public transportation’, ‘tobacco products’, ‘apparel’, ‘health goods and services’ and ‘reading materials’.

The Cobb–Douglas representative agent bias turns out to be 0.19% in both scenarios. While the bias is higher with a more disaggregated classification, it still remains small.

5. Representative Agent Bias in the Tornqvist Index

Computing the bias in the Tornqvist index requires panel data at the household level to get information about the base and current period budget shares. Unfortunately, panel data on commodity-specific detailed consumption expenditure is not very common. The Panel Study of Income Dynamics (PSID) collects longitudinal annual data on households in the United States. A limitation of this panel is that it collects data only for a subset of consumption items, mainly food. However, Blundell *et al*. (2008) combine the longitudinal PSID and the cross-sectional CEX to impute consumption for other categories.

The imputation procedure is implemented as follows. All the CEX data from 1980 to 1992 is pooled, and for any individual *j* in period *t*, the demand equation for food is written as

(3) $f_{j.t} = \boldsymbol{W}'_{\boldsymbol{j.t}}\boldsymbol{\mu} + \boldsymbol{p}'_{\boldsymbol{j}}\boldsymbol{\theta} + \beta(D_{j.t})c_{j.t} + e_{j.t}$

where *f* is the log of real food expenditure (available in both surveys), W and $\boldsymbol{p}$ contain a set of demographic variables and relative prices (also available in both datasets), *c* is the log of non-durable expenditure (available only in the CEX) and *e* captures unobserved heterogeneity in the demand for food and measurement error in food expenditure. The elasticity β (budget elasticity) varies with time and with observable household characteristics (*D*). To account for measurement error of total non-durable expenditure, Blundell *et al*. (2008) instrument the latter with the average of the hourly wage of the husband and the average of the hourly wage of the wife (both are averaged by cohort, year and education). Armed with these estimates, the non-durable consumption for all households in the PSID is estimated by inverting the demand function. A similar procedure is used to impute total expenditure and hence calculate expenditure on durables.

We draw on this dataset created by Blundell *et al*. (2008) for the period 1980–92 and construct a Tornqvist index for three commodity categories: food, non-durables and ‘others’, which includes durables, health and education. Figures 1–3 show the kernel densities of budget shares (for various consumption categories) for some of these years. Just as in the Indian data, there is considerable heterogeneity in budget shares in the US data as well.

The Blundell *et al*. (2008) dataset also contains yearly price indices (from the BLS) for food and for non-durables. The indices are US city averages, not seasonally adjusted, and are expressed in 1982–84 dollars. Since the budget shares for all components are known, we derive the price index for ‘other commodities’ (which includes durables, health and education) from the aggregate CPI and the price indices for food and non-durables.

Table 7 displays the change in relative prices between the base year 1980 and some of the other years. Table 7 also shows the variation in budget shares of the three consumption categories. The resulting representative agent bias is presented in the last row. It is very small.

6. An Upper Bound to the Bias in the Tornqvist Index

As noted earlier, panel data on consumption expenditure is not commonly available. In the United States, for instance, the CEX provides a comprehensive dataset on the spending habits of US households, but it follows households for only four quarters at most. While a quarterly rotating panel can be constructed with this data, it does not capture the variation across time periods adequately. Other panel datasets widely used by economists, such as the National Longitudinal Survey (NLS) or the Health and Retirement Survey (HRS), have abundant information on income or wealth, but no information whatsoever on consumption. In the UK, the Family Expenditure Survey (FES) provides comprehensive data on household expenditures, but households are not followed over time. Panel datasets that collect data on income or wealth, such as the British Household Panel Survey (BHPS), typically lack consumption data.

While the representative agent bias in the Tornqvist index cannot be estimated from cross-sectional data, in this section we provide an upper bound to the bias that can be estimated by cross-sectional data alone.

From (1), the representative agent bias for Tornqvist index can be expressed as

(4) $g \approx (\frac{1}{2!}) \sum_{i=1}^{M-1} var(s_i) (\ln \lambda_i)^2 + (\frac{1}{2!}) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} cov(s_i s_k) (\ln \lambda_i)(\ln \lambda_k)$

Both the variance and the covariance terms above require household budget share data for a base and a current period. However, we show in the appendix that

$$var(s_i) \leq \left(\frac{1}{4}\right) \left[var(s_i^1) + var(s_i^0) + 2\sqrt{var(s_i^1) var\left(s_i^0\right)} \right] \forall i = 1,2, \ldots, M-1$$

where the right-hand side can now be computed by cross-sectional data for the base and current periods. Further, it is also shown that if the change in budget share for the *i*th commodity is independent of the budget share of the *k*th commodity in the base period and similarly the change in budget share for the *k*th commodity is independent of the budget share of the *i*th commodity in the base period, then the covariance terms can be approximated by

$$cov\left(s_{i,} s_k\right) = \left(\frac{3}{4}\right) cov(s_i^0, s_k^0) + (\frac{1}{4}) cov(s_i^1, s_k^1)$$

or they can also be approximated by

$$cov(s_i, s_k) = \left(\frac{1}{4}\right) cov(s_i^0, s_k^0) + (\frac{3}{4}) cov(s_i^1, s_k^1)$$

As the covariances of budget shares are very small (they approximate to zero when rounded to the second decimal place), it is of little consequence which approximation is used. If the first approximation is used, an upper bound to the representative agent bias is derived as

$$(5)\ g \leq \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \left[var(s_i^1) + var(s_i^0) + 2\sqrt{var(s_i^1) var\left(s_i^0\right)} \right] (\ln \lambda_i)^2$$

$$+ \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} [\left(\frac{3}{8}\right) cov(s_i^0, s_k^0) + \left(\frac{1}{8}\right) cov(s_i^1, s_k^1)] (\ln \lambda_i)(\ln \lambda_k)$$

The right-hand side of expression (5) can be solely computed from cross-sectional data in the base and current periods.

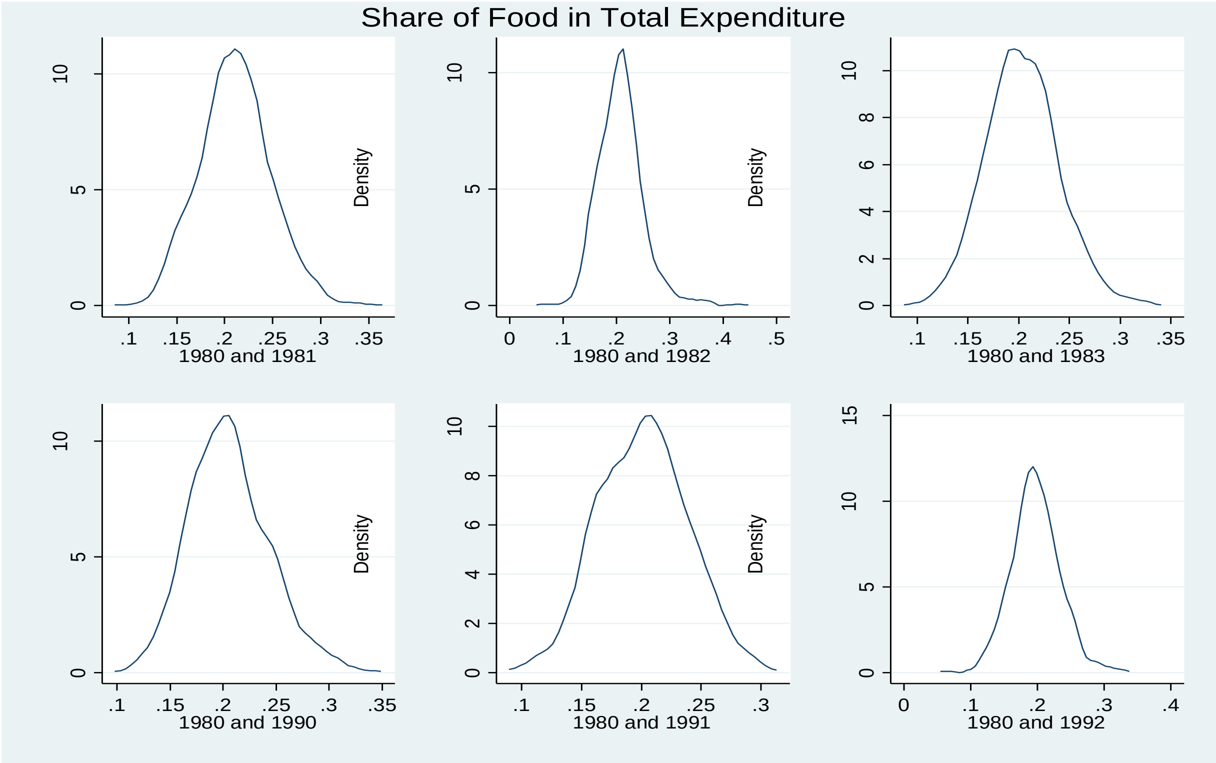
Table 8 displays the estimates of the upper bound to the representative agent bias in the Tornqvist index. All of these estimates are computed for the observed price changes. Columns 1 and 2 are the estimates from the cross-sectional Indian expenditure survey for rural and urban sectors. Column 3 is the estimate from the US CEX. Column 4 is the estimate from the PSID-based dataset of Blundell *et al*. (2008). In all cases, the upper bound turns out to be very small. Indeed, they turn out to be quite similar in magnitude to the bias in the Cobb–Douglas price index. Comparing the upper bound to the exact bias (for the US data set of Blundell *et al*.) shows the upper bound to be about twice the exact bias for the price change between1980 and 1992.

7. Concluding Remarks

It is well known that large changes in relative prices lead to substitution bias in the measurement of cost of living differences, and superlative indices have been devised as a way to minimize the bias. Even so, what this paper has shown is that the average of individual superlative COLIs is sensitive to heterogeneity in consumer spending patterns, whether because of variation in preferences or income. Conceptually, this means that the group COLI (which is what we are frequently called upon to interpret) depends not just on the change in prices or the levels of budget shares in the population, but also on the diversity of spending patterns in the population. The insight is significant in a practical sense, because statistical agencies do not usually calculate group COLIs. What they do is to evaluate the COLI at the average budget share. The resulting bias has been the focus of this paper.

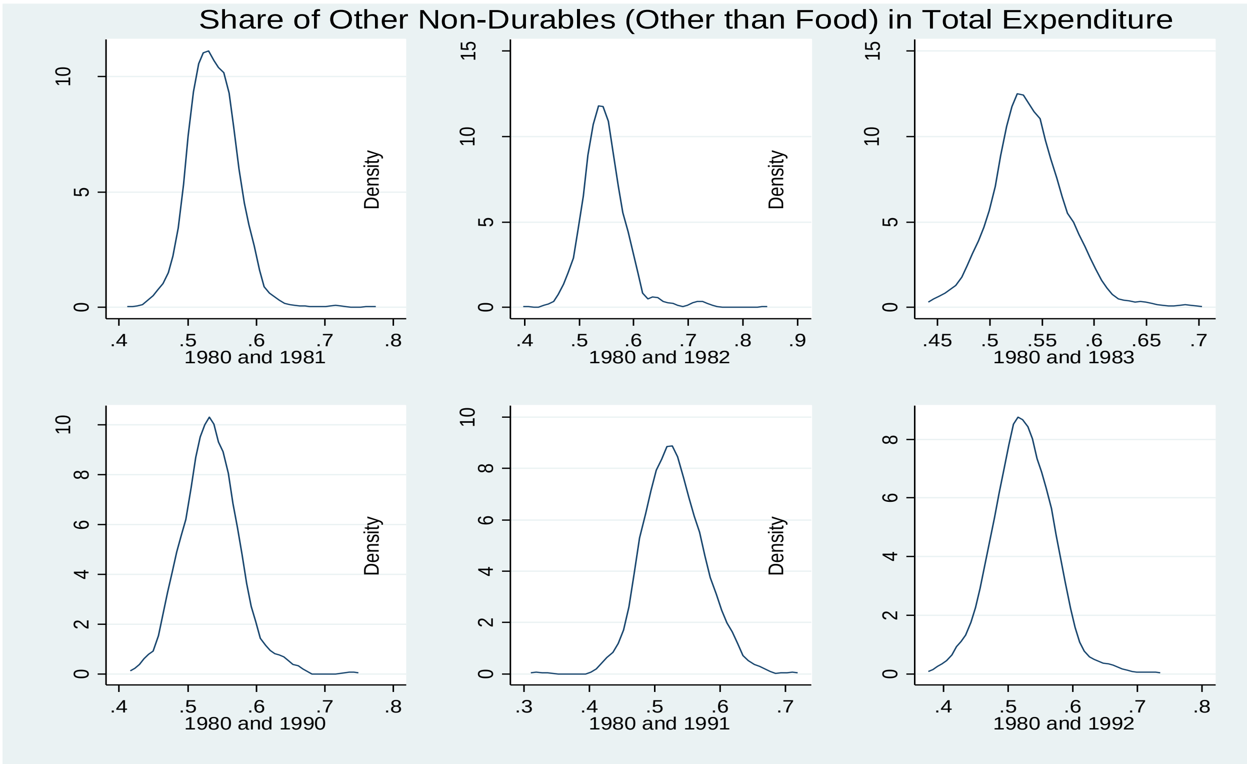
What this paper has shown is that for an important and widely used superlative index like the Tornqvist, the nature of the bias will be to underestimate the true group COLI. A similar result holds for the COLI generated from Cobb–Douglas preferences, which is widely used in applied welfare analysis. Furthermore, the magnitude of the bias depends on the extent to which the relative price structure changes between the base and current periods. The empirical exercises done in this paper for India and the United States show that the magnitude of the bias is, however, very small. It implies that the ‘representative agent bias’ may not be a serious problem for statistical agencies and researchers in price indices.

Figure 1: Kernel Density Estimate of Share of Food in Total Expenditure



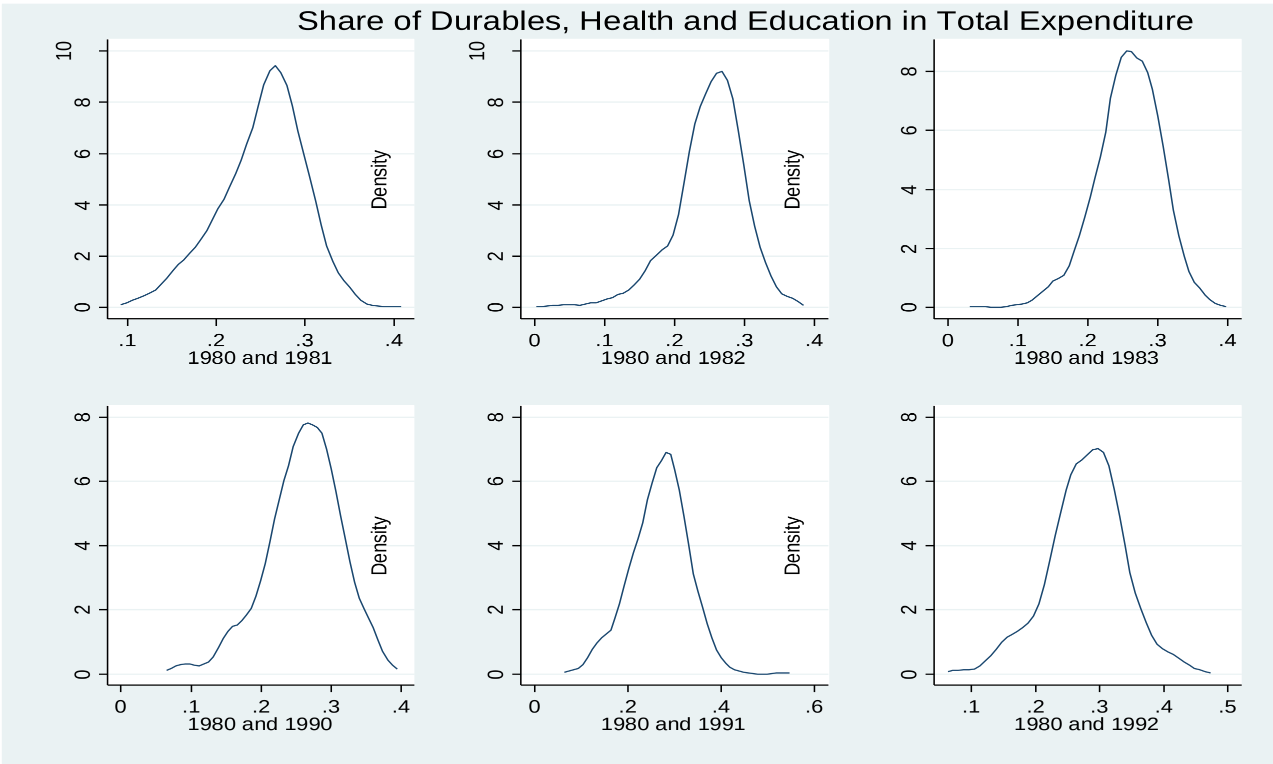
Note: Kernel density estimate of share of food in total expenditure for the years 1980–1981, 1980–1982,1980–1983,1980–1990, 1980–1991 and 1980–1992 (using the dataset created by Blundell *et al*., 2008).

Figure 2: Kernel Density Estimate of Share of Other Non-Durables in Total Expenditure



Note: Kernel density estimate of share of other non-durables in total expenditure for the years 1980–1981, 1980–1982,1980–1983,1980–1990, 1980–1991 and 1980–1992 (using the dataset created by Blundell *et al*., 2008).

Figure 3: Kernel Density Estimate of Share of Durables, Health and Education in Total Expenditure



Note: Kernel density estimate of share of durables, health and education in total expenditure for the years 1980–1981, 1980–1982,1980–1983,1980–1990,1980–1991 and 1980–1992 (using the dataset created by Blundell *et al*., 2008).

Table 1: Budget Share of Commodities (Indian Data: Rural)

Commodity	**Mean**	**Std. Dev.**	**CV(%)**
Cereals and cereal substitutes	0.2	0.09	45
Pulse and pulse products	0.03	0.02	67
Milk and milk products	0.07	0.08	114
Edible oil, fruits, fish and meat	0.09	0.04	44
Vegetables	0.07	0.03	43
Sugar, salt and spices	0.04	0.02	50
Beverages, tobacco and intoxicants	0.07	0.06	86
Fuel and light	0.1	0.04	40
Clothing	0.07	0.03	43
Bedding and footwear	0.04	0.09	225
Miscellaneous non-food	0.22	0.12	55

Note: Authors' calculation from National Sample Survey (2004–05) data. CV, coefficient of variation.

Table 2: Budget Share of Commodities (Indian Data: Urban)

Commodity	**Mean**	**Std. Dev.**	**CV(%)**
Cereals and cereal substitutes	0.13	0.07	54
Pulse and pulse products	0.03	0.01	33
Milk and milk products	0.08	0.05	62.5
Edible oil, fruits, fish and meat	0.08	0.04	50
Vegetables	0.05	0.03	60
Sugar, salt and spices	0.03	0.02	67
Beverages, tobacco and intoxicants	0.07	0.07	100
Fuel and light	0.1	0.04	40
Clothing	0.06	0.03	50
Bedding and footwear	0.04	0.09	225
Miscellaneous non-food	0.33	0.15	45

Note: Authors' calculation from National Sample Survey (2004–05) data. CV, coefficient of variation.

Table 3: Change in Prices and Representative Agent Bias (Indian Data: Rural)

Change in Prices (in %)	Scenario1	Scenario2	Scenario3
Cereals and cereal substitutes	65	80	20
Pulse and pulse products	109	80	20
Milk and milk products	115	20	80
Edible oil, fruits, fish and meat	15	20	80
Vegetables	95	20	80
Sugar, salt and spices	151	20	80
Beverages, tobacco and intoxicants	110	20	80
Fuel and light	101	20	80
Clothing	68	20	80
Bedding and footwear	68	20	80
Miscellaneous non-food	64	20	80
Representative agent bias (in %)	0.06	0.07	0.07

Note:In scenario 1, we consider the actual change in prices for all categories between 2011–12 and 2004–05. The change in relative prices is computed by deflating the change in the prices for all categories by the change in prices for miscellaneous non-food items, which we consider as the numeraire category. In scenario 2, we consider two different rates of change in relative prices. The prices of the most frequently consumed commodities by the poor (cereals and cereal substitutes; pulse and pulse products) are assumed to increase at a rate of 80%. Prices of other categories are assumed to increase at a rate of 20%. Scenario 3 is the exact opposite of scenario 2, where the prices of the most frequently consumed goods by the poor increase at a rate of 20% and the prices of other commodity group increase at a rate of 80%. All the reported figures are in percentages.

Table 4: Change in Prices and Representative Agent Bias (Indian Data: Urban)

Change in Prices (in %)	Scenario1	Scenario2	Scenario3
Cereals and cereal substitutes	73	80	20
Pulse and pulse products	107	80	20
Milk and milk products	107	20	80
Edible oil, fruits, fish and meat	32	20	80
Vegetables	89	20	80
Sugar, salt and spices	158	20	80
Beverages, tobacco and intoxicants	83	20	80
Fuel and light	55	20	80
Clothing	46	20	80
Bedding and footwear	46	20	80
Miscellaneous non-food	47	20	80
Representative agent bias (in %)	0.05	0.06	0.06

Note:In scenario 1, we consider the actual change in prices for all categories between 2011–12 and 2004–05. The change in relative prices is computed by deflating the change in prices for all categories by the change in prices for miscellaneous non-food items, which we consider as the numeraire category. In scenario 2, we consider two different rates of change in relative prices. The prices of the most frequently consumed commodities by the poor (cereals and cereal substitutes; pulse and pulse products) are assumed to increase at a rate of 80%. The prices of other categories are assumed to increase at a rate of 20%. Scenario 3 is the exact opposite of scenario 2, where the prices of the most frequently consumed goods by the poor increase at a rate of 20% and the prices of other commodity groups increase at a rate of 80%. All the reported figures are in percentages.

Table 5: Budget Share of Commodities (US Consumer Expenditure Data: 1982)

Commodities	**Mean**	**Std. Dev.**	**CV(%)**
Food consumed at home	0.14	0.11	79
Food consumed away from home	0.06	0.04	67
Other non-durable goods	0.57	0.12	21
Other goods (including durables, health and education)	0.23	0.14	61

Note:Authors' calculation from Consumer Expenditure Survey data in the United States (using the dataset created by Blundell*et al*., 2008).

Table 6: Change in Prices and Representative Agent Bias for Cobb–Douglas Price Index (US Data, between 1982 and 1992)

Change in Prices (in%)	**Scenario1**	**Scenario2**	**Scenario3**
Food consumed at home	39	80	20
Food consumed away from home	47	80	20
Other non-durable goods	30	20	80
Other goods (including durables, health and education)	65	20	80
Representative agent bias (in %)	0.046	0.06	0.06

Note:Authors' calculation from Consumer Expenditure Survey data in the United States (using the dataset created by Blundell*et al*.,2008). In scenario 1, we consider the actual change in prices for all categories between 1982 and 1992. The change in relative prices is computed by deflating the change in the prices for all categories by the change in prices for other goods (including durables, health and education), which we consider as the numeraire category. In scenario 2, we consider two different rates of change in relative prices. The prices of food (i.e. for the categories 'food consumed at home' and 'food consumed away') are assumed to increase at a rate of 80%. The prices of other categories are assumed to increase at a rate of 20%. Scenario 3 is the exact opposite of scenario 2, where the prices of 'food consumed at home' and 'food consumed away' increase at the a of 20% and the prices of other commodity groups increase at a rate of 80%. All the reported figures are in percentages.

Table 7: Representative Agent Bias for Tornqvist Index (3 commodities: ‘food’, ‘non-durables other than food’ and ‘other commodities’, which includes durables, health and education)

Year	Between 1981 and 1980	Between 1982 and 1980	Between 1983 and 1980	Between 1990 and 1980	Between 1991 and 1980	Between 1992 and 1980
Change in relative price of ‘food’ with respect to ‘other commodities’ (other than non-durables)	8.3	18.1	21.6	26.1	27.1	29.6
Change in relative price of ‘other non-durables’ (other than food) with respect to ‘other commodities’ (other than non-durables)	7	18.8	22.5	34.3	34.6	36.1
Coefficient of variation for the budget share of food (in %)	19	19	20	19	20	20
Coefficient of variation for the budget share of other non-durables (in %)	7.4	7	7.4	7.5	9	10
Coefficient of variation for the budget share of other commodities in total expenditure (in %)	20	25	19	19	22	21
Bias (in %)	0.0009	0.008	0.007	0.019	0.027	0.031

Note: Authors’ calculation based on Panel Study of Income Dynamics and Consumer Expenditure Survey data in the United States (using the dataset by Blundell *et al*., 2008). All the reported figures are in percentages.

Table 8: Upper Bound to the Representative Agent Bias in the Tornqvist Index

	India (Rural)	India (Urban)	United States (Consumer Expenditure Survey)	United States (Panel Study of Income Dynamics–Based Dataset, between 1992 and 1980)
Upper Bound (in %)	0.06	0.05	0.06	0.06

Note: Authors'calculations. For India, the calculations are based on the National Sample Survey (2004–05). For the United States, the calculations are based on Panel Study of Income Dynamics and Consumer Expenditure Survey data (using the data set by Blundell *et al*., 2008).

Appendix

1. Proof of Proposition1

Convexity of $T(\boldsymbol{s})$requires the matrix of the second derivative of $T(\boldsymbol{s})$**,** i.e. the Hessian matrix, to be positive semidefinite. A diagonal element of the matrix is $\frac{\partial^2 T}{\partial {s_i}^2} = T.[ln\lambda_i^2]\ \forall i = 1,2, \dots, M - 1$, where$T$ is the Tornqvist index. An off-diagonal element can be written as $\frac{\partial^2 T}{\partial s_i \partial s_k} = T.[ln\lambda_i ln\lambda_k]\ \ \forall i = 1,2, \dots, M-1, k = 1,2, \dots, M-1; i \neq k.$

Hence the Hessian matrix can be written as

$$H = T(D.D^t)$$

where $D^t$is the $M - 1$ row vector of $(ln\lambda_1, ln\lambda_2, \dots, ln\lambda_{M-1})$ and D is its transpose. For every non-zero column vector Y belonging to the M-1 dimensional real space, we can write $Y^t HY = Y^t\ T(D.D^t)Y = T(Y^t D.D^t Y) = T\big((D^t Y)^t (D^t Y)\big) = T||D^t Y\ ||^2 \geq 0$

Hence $T(\boldsymbol{s})$ is convex in the vector budget shares, i.e. $\boldsymbol{s}$.

2. Proof of Proposition3

Considering a second-order Taylor's series expansion of $T(\boldsymbol{s})$ around $(\boldsymbol{s})$, we obtain

$$T(\boldsymbol{s}) = T[\boldsymbol{E}(\boldsymbol{s})] + \sum_{i=1}^{M-1}\big(s_i - E(s_i)\big)\left(\frac{\partial T}{\partial s_i}\right) + \left(\frac{1}{2!}\right)\sum_{i=1}^{M-1}\big(s_i - E(s_i)\big)^2\left(\frac{\partial^2 T}{\partial s_i^2}\right)$$

(A1) $$+\left(\frac{1}{2!}\right)\sum_{i=1}^{M-1}\sum_{k=1;i\neq k}^{M-1}\big(s_i - E(s_i)\big)\big(s_k - E(s_k)\big)\left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right) + R_2$$

R_2 is the remainder term corresponding to the second-order Taylor's series approximation in equation (A1). Let $h_i = s_i - E(s_i)$ and $\boldsymbol{h}$ be the vector $(h_1 h_2 \ldots h_{M-1})$. Let

$$||\boldsymbol{h}|| = \sqrt{(h_1^2 + h_2^2 + \cdots h_{M-1}^2)}$$

It can be shown that $R_2(\boldsymbol{E}(\boldsymbol{s}), \boldsymbol{h})$ is $o(||\boldsymbol{h}||^2)$, i.e. $\frac{R_2(\boldsymbol{E}(\boldsymbol{s}),\boldsymbol{h})}{||\boldsymbol{h}||^2}$ tends to zero as $\boldsymbol{h}$ tends to zero (the details about the remainder term are discussed later in the appendix).

Taking expectation on both sides of equation (A1) and rearranging, we get

$$E[T(\boldsymbol{s})] - T[\boldsymbol{E}(\boldsymbol{s})] \approx \left(\frac{1}{2!}\right) \sum_{\mathrm{i}=1}^{\mathrm{M}-1} \mathrm{E}[\, s_i - E(s_i)]^2 \left(\frac{\partial^2 T}{\partial s_i^2}\right)$$

$$+\left(\frac{1}{2!}\right) \sum_{\mathrm{i}=1}^{\mathrm{M}-1} \sum_{\mathrm{k}=1,\mathrm{i}\neq\mathrm{k}}^{\mathrm{M}-1} \mathrm{E}\big[\big(s_i - E(s_i)\big)\big(s_k - E(s_k)\big)\big] \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)$$

(A2) $$= \left(\frac{1}{2!}\right) [\textstyle\sum_{i=1}^{M-1} var(s_i) \left(\frac{\partial^2 T}{\partial s_i^2}\right) + \sum_{\mathrm{i}=1}^{\mathrm{M}-1} \sum_{\mathrm{k}=1,\mathrm{i}\neq\mathrm{k}}^{\mathrm{M}-1} \mathrm{cov}(s_i, s_k) \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)$$

Dividing both sides of equation (A2) by $T[E(\boldsymbol{s})]$, we get

$$\frac{E[T(\boldsymbol{s})] - T[\boldsymbol{E}(\boldsymbol{s})]}{T[\boldsymbol{E}(\boldsymbol{s})]} \approx \left(\frac{1}{2!}\right) \sum_{i=1}^{M-1} var(\, s_i) \frac{\left(\frac{\partial^2 T}{\partial s_i^2}\right)}{T[\mathrm{E}(\boldsymbol{s})]}$$

(A3) $$+\left(\frac{1}{2!}\right) \textstyle\sum_{i=1}^{M-1} \sum_{k=1,i\neq k}^{M-1} cov(\, s_i, s_k) \frac{\left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)}{T[\mathrm{E}(\boldsymbol{s})]}$$

Now, $\frac{\left(\frac{\partial^2 T}{\partial s_i^2}\right)}{T[\boldsymbol{E}(\boldsymbol{s})]} = (\ln \lambda_i)^2 \;\forall\; i = 1,2, \ldots, M-1$ and $\frac{\frac{\partial^2 T}{\partial s_i \partial s_k}}{T[\boldsymbol{E}(\boldsymbol{s})]} = (\ln \lambda_i)(\ln \lambda_k)] \;\forall\; i = 1,2, \ldots, M-1;\; k = 1,2, \ldots, M-1;\; i \neq \mathrm{k}$

Plugging these values in equation (A3), we obtain the following:

$$g \approx (\frac{1}{2!})\sum_{i=1}^{M-1} var(s_i)(\ln \lambda_i)^2 + (\frac{1}{2!})\sum_{i=1}^{M-1}\sum_{k=1, i\neq k}^{M-1} cov(s_i, s_k)(\ln \lambda_i)(\ln \lambda_k)$$

Therefore, the representative agent bias is characterized by

(A4) $$g \approx (\frac{1}{2!}) var[\sum_{i=1}^{M-1} s_i \ln \lambda_i]$$

The expression (A4) is the same as equation (1), as shown in the main text.

Returning to the remainder term, it can be represented in different forms. The following result is based on a version of the Lagrange form. If there exists a positive constant U, such that

$$|(\frac{\partial}{\partial s_1} + \frac{\partial}{\partial s_2} + \cdots + \frac{\partial}{\partial s_{M-1}})^3 T[\boldsymbol{t}]| \leq U$$

$\forall \boldsymbol{t} = (t_1 t_2 \ldots t_{M-1})$; $t_i \epsilon [\mathrm{E}(s_i), \mathrm{E}(s_i) + \mathrm{h}_i]$ when h_i is positive and $\boldsymbol{t_i} \epsilon [\mathrm{E}(s_i) + \mathrm{h}_i, \mathrm{E}(s_i)]$, when $\mathrm{h}_i$is negative ($\forall i = 1,2, \ldots, M-1$), then the remainder term can be bounded as

$$R_2(\boldsymbol{E}(\boldsymbol{s}), \boldsymbol{h}) \leq \frac{||\boldsymbol{h}||^3}{3!} \mathrm{U}$$

It can be readily checked that $\frac{||\boldsymbol{h}||^3}{3!}\mathrm{U}$ is $o(||\boldsymbol{h}||^2)$, i.e. dividing $\frac{||\boldsymbol{h}||^3}{3!}\mathrm{U}$ by $||\boldsymbol{h}||^2$, we get $\frac{||\boldsymbol{h}||}{3!}\mathrm{U}$and this goes to zero as$\boldsymbol{h} \rightarrow 0$(provided thatU is a positive constant). As $\frac{||\boldsymbol{h}||^3}{3!}\mathrm{U}$ is $o(||\boldsymbol{h}||^2)$and $R_2(\boldsymbol{E}(\boldsymbol{s}), \boldsymbol{h}) \leq \frac{||\boldsymbol{h}||^3}{3!}\mathrm{U}$, $R_2(\boldsymbol{E}(\boldsymbol{s}), \boldsymbol{h})$ is $o(||\boldsymbol{h}||^2)$ as well, i.e.$\frac{R_2(\boldsymbol{E}(\boldsymbol{s}),\boldsymbol{h})}{||\boldsymbol{h}||^2}$tends to zero as $\boldsymbol{h}$ tends to zero.

The only thing we need to show is that U is a positive constant and U satisfies the following condition:

$$|(\frac{\partial}{\partial s_1} + \frac{\partial}{\partial s_2} + \cdots + \frac{\partial}{\partial s_{M-1}})^3 T[\boldsymbol{t}]| \leq U$$

Now,

$$|(\frac{\partial}{\partial s_1} + \frac{\partial}{\partial s_2} + \cdots + \frac{\partial}{\partial s_{M-1}})^3 T(\boldsymbol{t})|$$

$$= |\sum_{i=1}^{M-1} \frac{\partial^3 T(\boldsymbol{t})}{\partial s_i^3} + 3\sum_{i=1}^{M-1} \sum_{k=1;i\neq k}^{M-1} \frac{\partial^3 T(\boldsymbol{t})}{\partial s_i^2 \partial s_k}$$

$$+ \sum_{i=1}^{M-1} \sum_{k=1}^{M-1} \sum_{l=1;i\neq k\neq l}^{M-1} \frac{\partial^3 T(\boldsymbol{t})}{\partial s_i \partial s_k \partial s_l}$$

Since we are considering the absolute value of the derivative, $(\frac{\partial}{\partial s_1} + \frac{\partial}{\partial s_2} + \cdots + \frac{\partial}{\partial s_{M-1}})^3 T(\boldsymbol{t})$, it is always positive. As long as the third-order own and cross partial derivatives are finite, an upper bound U of the derivatives exists. Therefore, a positive constant U exists as an upper bound.

3. Derivation of the Upper Bound

The representative agent bias for the Tornqvist index can be expressed as

$$g \approx (\frac{1}{2!}) \sum_{i=1}^{M-1} var(s_i)(\ln \lambda_i)^2 + (\frac{1}{2!}) \sum_{i=1}^{M-1} \sum_{k=1,i\neq k}^{M-1} cov(s_i, s_k)(\ln \lambda_i)(\ln \lambda_k)$$

$$s_i = \left(\frac{1}{2}\right)(s_i^1 + s_i^0);\ s_k = \left(\frac{1}{2}\right)(s_k^1 + s_k^0)$$

The bias cannot be computed without panel data at the household level. But we can generate upper bounds on the bias, which can be computed from cross-sectional data. Suppose we split up the expression for representative agent bias into two parts. The first part of the bias is

$$A = (\frac{1}{2!})\sum_{i=1}^{M-1} var(s_i)\ (\ln \lambda_i)^2$$

Now,

$$var(s_i) = var\left(\frac{s_i^1 + s_i^0}{2}\right) = \left(\frac{1}{4}\right) var(s_i^1 + s_i^0)$$

$$= \left(\frac{1}{4}\right) [var(s_i^1) + var(s_i^0) + 2cov(s_i^1, s_i^0)]$$

The term $cov(s_i^1, s_i^0)$ cannot be computed because of the lack of panel data. But we can generate an upper bound on the expression of the variance, i.e. $var(s_i)$. In order to generate that upper bound, the expression of the variance is written in the following way:

$$var(s_i) = var\left(\frac{s_i^1 + s_i^0}{2}\right) = \left(\frac{1}{4}\right) var(s_i^1 + s_i^0)$$

$$= \left(\frac{1}{4}\right)\left[var(s_i^1) + var(s_i^0) + 2\frac{cov(s_i^1, s_i^0)}{\sqrt{var(s_i^1)var(s_i^0)}}\sqrt{var(s_i^1)var(s_i^0)}\right]$$

Now,

$$\frac{cov(s_i^1, s_i^0)}{\sqrt{var(s_i^1)var(s_i^0)}} = (R_i^2)^{(\frac{1}{2})}$$

whereR_i^2 is the squared correlation coefficient between s_i^1 and s_i^0 $\forall i = 1, 2, \dots, M-1$. Replacing $\frac{cov(s_i^1, s_i^0)}{\sqrt{var(s_i^1)var(s_i^0)}}$ by $(R_i^2)^{(\frac{1}{2})}$, we can write down the variance as

$$var(s_i) = \left(\frac{1}{4}\right)\left[var(s_i^1) + var(s_i^0) + 2(R_i^2)^{(\frac{1}{2})}\sqrt{var(s_i^1)var(s_i^0)}\right]$$

The maximum value of R_i^2 can be 1. Putting this maximum value of R_i^2 in the variance expression, we obtain the following upper bound on the variance:

$$var(s_i) = \left(\frac{1}{4}\right)\left[var(s_i^1) + var(s_i^0) + 2(R_i^2)^{(\frac{1}{2})}\sqrt{var(s_i^1)var(s_i^0)}\right]$$

$$\leq \left(\frac{1}{4}\right)\left[var(s_i^1) + var(s_i^0) + 2\sqrt{var(s_i^1)var(s_i^0)}\right] \forall i = 1,2, \dots, M-1$$

The imposition of an upper bound on the variance generates an upper bound on the first term of the bias expression, which we can write down as

$$A = \left(\frac{1}{2!}\right)\sum_{i=1}^{M-1} var(s_i)\ (\ln \lambda_i)^2$$

(A5) $$\leq \left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\left[var(s_i^1) + var(s_i^0) + 2\sqrt{var(s_i^1)var(s_i^0)}\right] (\ln \lambda_i)^2$$

Now we focus on the second term of the bias expression, which we can write as $B = (\frac{1}{2!})\sum_{i=1}^{M-1}\sum_{k=1, i\neq k}^{M-1} cov(s_i, s_k)(\ln \lambda_i)(\ln \lambda_k)$

The covariance term, i.e.$cov(s_{i,} s_k)$, can be further rewritten in the following way:

$cov(s_{i,}s_k) = cov\left(\left(\frac{1}{2}\right)(s_i^1 + s_i^0), \left(\frac{1}{2}\right)(s_k^1 + s_k^0)\right) = \left(\frac{1}{4}\right)[cov(s_i^1, s_k^1) + cov(s_i^0, s_k^0)] +$

$\left(\frac{1}{4}\right)[cov(s_i^1, s_k^0) + cov(s_i^0, s_k^1)]; \forall i = 1, 2, \ldots, M-1; k = 1, 2, \ldots, M-1; i \neq k$

The terms $cov(s_i^1, s_k^1)$ and $cov(s_i^0, s_k^0)$ can be directly computed from the cross-sectional data on budget shares corresponding to period 1 and period 0, respectively. But we cannot compute $cov(s_i^1, s_k^0)$ and $cov(s_i^0, s_k^1)$ without panel data. In order to compute these two terms, we simplify them further. We write down the current period (period 1) budget shares as$s_i^1 = s_i^0 + u_i$and $s_k^1 = s_k^0 + u_k$, where u_i and u_k are the changes in the budget share of the*i*th and *k*th commodity, respectively (between period 1 and period 0). Therefore we can write down $cov(s_i^1, s_k^0)$as

$$cov(s_i^1, s_k^0) = cov(s_i^0 + u_i, s_k^0) = cov(s_i^0, s_k^0) + cov(s_k^0, u_i)$$

If we assume that the change in the budget share of the*i*th commodity (between period 1 and period 0) is independent of the budget share of the *k*th commodity in the base period, then $cov(s_i^1, s_k^0)$ equals $cov(s_i^0, s_k^0)$. Similarly,

$$cov(s_i^0, s_k^1) = cov(s_i^0, s_k^0 + u_k) = cov(s_i^0, s_k^0) + cov(s_i^0, u_k)$$

Therefore,$cov(s_i^0, s_k^1) = cov(s_i^0, s_k^0)$ under the assumption that the budget share of the *i*th commodity in the base period is independent of the change in the budget share of the *k*th commodity (between period 1 and period 0). Hence we can write down$cov(s_i, s_k)$ as

$$cov(s_{i,}s_k) = \left(\frac{3}{4}\right)cov(s_i^0, s_k^0) + (\frac{1}{4})cov(s_i^1, s_k^1)$$

Therefore,.$cov(s_i, s_k)$ can be approximated from cross-sectional data on budget shares at the household level (for period 1 and period 0) using the above expression. Under this approximation, the second term of the bias (which we denote as B) becomes

$$\left(\frac{1}{2!}\right)\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}[\left(\frac{3}{4}\right)cov(s_i^0,s_k^0)+\left(\frac{1}{4}\right)cov(s_i^1,s_k^1)](\ln\lambda_i)(\ln\lambda_k)$$

(A6) $$=\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}[\left(\frac{3}{8}\right)cov(s_i^0,s_k^0)+\left(\frac{1}{8}\right)cov(s_i^1,s_k^1)](\ln\lambda_i)(\ln\lambda_k)$$

Combining the upper bound on the first term of the bias (A5) and the approximation of the second term of the bias (A6), we write down the overall upper bound of the representative agent bias as

$$g\approx(\frac{1}{2!})\sum_{i=1}^{M-1}var(s_i\,(\ln\lambda_i)^2)+(\frac{1}{2!})\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}cov(s_i,s_k)\,(\ln\lambda_i)(\ln\lambda_k)$$

$$\leq\left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\left[var(s_i^1)+var(s_i^0)+2\sqrt{var(s_i^1)var(s_i^0)}\right](\ln\lambda_i)^2$$

$$+\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}[\left(\frac{3}{8}\right)cov(s_i^0,s_k^0)+\left(\frac{1}{8}\right)cov(s_i^1,s_k^1)](\ln\lambda_i)(\ln\lambda_k)$$

There is an alternative way to approximate $cov(s_i^1,s_k^0)$ and $cov(s_i^0,s_k^1)$. We can write down $cov(s_i^1,s_k^0)=cov(s_i^1,s_k^1-u_k)=cov(s_i^1,s_k^1)$ if $cov(s_i^1,u_k)=0$, i.e. the change in the budget share of the *k*th commodity (between period 1 and period 0) is independent of the budget share of the *i*th commodity in the current period (period 1). Similarly, $cov(s_k^1,u_i)=0$ implies $cov(s_i^0,s_k^1)=cov(s_i^1,s_k^1)$. Under this alternative approach, $cov(s_i,s_k)$ approximately equals $\left(\frac{1}{4}\right)cov(s_i^0,s_k^0)+(\frac{3}{4})cov(s_i^1,s_k^1)$ and hence the upper bound of the representative agent bias becomes

$$g\approx(\frac{1}{2!})\sum_{i=1}^{M-1}var(s_i\,(\ln\lambda_i)^2)+(\frac{1}{2!})\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}cov(s_i,s_k)\,(\ln\lambda_i)(\ln\lambda_k)$$

$$\leq \left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\left[var(s_i^1) + var(s_i^0) + 2\sqrt{var(s_i^1)var\left(s_i^0\right)}\right](\ln \lambda_i)^2$$

$$+\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}[\left(\frac{1}{8}\right)cov(s_i^0, s_k^0) + \left(\frac{3}{8}\right)cov(s_i^1, s_k^1)](\ln \lambda_i)(\ln \lambda_k)$$

These two alternative expressions of upper bounds turn out to be identical empirically (mentioned in section 6 of this paper).

References

Abraham, K. G., Greenlees, J. S. and Moulton, B. R. (1998). 'Working to improve the consumer price index',*Journal of Economic Perspectives*, **12**(1), pp. 27–36.

Almås, I. and Kjelsrud, A. (2017). 'Rags and riches: Relative prices, non-homothetic preferences, and inequality in India',*World Development*, **97**, pp. 102–21.

Blundell, R., Pistaferri, L. and Preston, I. (2008). 'Consumption inequality and partial insurance', *American Economic Review*, **98**(5), pp. 1887–921.

Boskin, M. J., Dulberger, E. L., Gordon, R. J., Griliches, Z. and Jorgenson, D. W. (1998). 'Consumer prices, the consumer price index, and the cost of living', *Journal of Economic Perspectives*, **12**(1), pp. 3–26.

Cage, R. A., Garner, T. and Ruiz-Castillo, J. (2002). 'Constructing household specific consumer price indexes: An analysis of different techniques and methods'. BLS Working Paper No. 354,March.

Caves, D. W., Christensen, L. R. and Diewert, W. E. (1982). 'The economic theory of index numbers and the measurement of input, output, and productivity', *Econometrica: Journal of the Econometric Society*, 50(6), pp. 1393–1414.

Crawford, I. (1994). 'UK household cost-of-living indices, 1979–92', *Fiscal Studies*, **15**(4), pp. 1–28.

Crawford, I. and Smith, Z. (2002). 'Distributional aspects of inflation, commentary no. 90'. Institute for Fiscal Studies, London.

Crossley, T. F. and Pendakur, K. (2010). 'The common-scaling social cost-of-living index', *Journal of Business & Economic Statistics*, **28**(4), pp. 523–38.

Deaton, A., & Muellbauer, J. (1980). *Economics and consumer behavior*. Cambridge university press.

Del Río, C. and Ruiz-Castillo, J. (2002).'Demographic trends and living standards: The case of Spain during the 1980s', *Revista de EconomíaAplicada*, **X**,pp. 217–34.

Diewert, W. E. (1976). 'Exact and superlative index numbers',*Journal of Econometrics*, **4**(2), pp. 115–45.

Diewert, W. E. (1978). 'Superlative index numbers and consistency in aggregation',*Econometrica: Journal of the Econometric Society*, 46(4), pp. 883–900.

Dumagan, J. C. (2002). 'Comparing the superlative Törnqvist and Fisher ideal indexes',*Economics Letters*, **76**(2), pp. 251–8.

Fisher, F. M. and Griliches, Z. (1995). 'Aggregate price indices, new goods, and generics', *Quarterly Journal of Economics*, **110**(1), pp. 229–44.

Garner, Thesia I., David S. Johnson, and MaryF. Kokoski,''An Experimental Consumer PriceIndex for the Poor,''*Monthly Labor Review*, September 1996,119:9, 32–42

Garner, T. I., Ruiz-Castillo, J. and Sastre, M. (2003). 'The influence of demographics and household-specific price indices on consumption-based inequality and welfare: A comparison of Spain and the United States',*Southern Economic Journal*, 70(1), pp. 22–48.

Hagemann, R. P. (1982). 'The variability of inflation rates across household types',*Journal of Money, Credit and Banking*, **14**(4), pp. 494–510.

Hobijn, B. and Lagakos, D. (2003). 'Social security and the consumer price index for the elderly', *Current Issues in Economics and Finance*, **9**(5).

Idson, T. and Miller, C. (1994).'The effects of household characteristics on household-specific inflation rates: An application to trends in child poverty and educational real wage differentials'. Discussion paper 695, Columbia University, Dept. of Economics.

Johnson, D. S., Parker, J. A. and Souleles, N. S. (2006).'Household expenditure and the income tax rebates of 2001',*American Economic Review*,**96**(5),pp. 1589–610.

Kaplan, G. andSchulhofer-Wohl, S. (2017). 'Inflation at the household level',*Journal of Monetary Economics*, **91**, pp. 19–38.

Kokoski, M. F. (1987). 'Problems in the measurement of consumer cost-of-living indexes',*Journal of Business & Economic Statistics*, **5**(1), pp. 39–46.

Livada, A. (1990). 'The distribution of household inflation rates: The Greek experience',*Bulletin of Economic Research*, **42**(3), pp. 175–96.

Lyssiotou, P. and Pashardes, P. (2004).'Comparing the true cost of living indices of demographically different households',*Bulletin of Economic Research*,**56**(1),pp. 21–39.

Mackie, C. and Schultze, C.L. (2002).*At What Price? Conceptualizing andMeasuring Cost-of-Living Indexes*. Washington, DC: National Academy Press.

Manser, M. E. and McDonald, R. J. (1988). ‘An analysis of substitution bias in measuring inflation, 1959–85’, *Econometrica: Journal of the Econometric Society*, pp. 909–930.

Michael, R. T. (1979). ‘Variation across households in the rate of inflation’, *Journal of Money, Credit and Banking*, **11**(1), pp. 32–46.

Muellbauer, J. (1974). ‘The political economy of price indices’. Birkbeck Discussion Paper no. 22, March.

Nicholson, J. L. (1975). ‘Whose cost of living?’ *Journal of the Royal Statistical Society. Series A (General)*, **138**(4), pp. 540–2.

Pollak, R. A. (1980). ‘Group cost-of-living indexes’,*American Economic Review*, **70**(2), pp. 273–8.

Pollak, R. A. (1981). ‘The social cost of living index’, *Journal of Public Economics*, **15**(3), pp. 311–36.

Porto, G. G. (2006). ‘Using survey data to assess the distributional effects of trade policy’,*Journal of International Economics*, **70**(1), pp. 140–60.

Prais, S. J. (1959). ‘Whose cost of living?’ *Review of Economic Studies*, **26**(2), pp. 126–34.

Prud’homme, M., Sanga, D. and Yu, K. (2005). ‘A computer software price index using scanner data’, *Canadian Journal of Economics/Revue Canadienne d'économique*, **38**(3), pp. 999–1017.

Rothschild, M. and Stiglitz, J. E. (1970). ‘Increasing risk: I. A definition’,*Journal of Economic Theory*, **2**(3), pp. 225–43.

Stewart, K. J. (2008). ‘The experimental consumer price index for elderly Americans (CPI-E): 1982–2007’,*Monthly Labor Review*, **131 (4)**, p. 19-24