Received: 4 December 2019 | Revised: 22 January 2021 | Accepted: 6 February 2021

DOI: 10.1111/boer.12286

BULLETIN OF ECONOMIC RESEARCH

ARTICLE

The representative agent bias in cost of living indices

Sutirtha Bandyopadhyay[1] | **Bharat Ramaswami**[2]

[1] Economics Area, Indian Institute of Management Indore, Indore, India

[2] Economics Department, Ashoka UniversityRajiv Gandhi Education City, Sonepat, India

Correspondence
Sutirtha Bandyopadhyay, Economics Area, Indian Institute of Management Indore, Indore, 453 556, India.
Email: sutirtha.bandyopadhyay@gmail.com

Abstract

The aggregate cost of the 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 difference between the two consists of a "plutocratic" bias and a "curvature" bias. The former is well known, but the latter has not been recognized earlier. In empirical applications, the curvature bias is small and orders of magnitude smaller than the plutocratic bias. This suggests much of the overall bias would be removed by constructing the representative agent by democratic averages of budget shares.

KEYWORDS

aggregation, cost of living index, curvature bias, heterogeneity, plutocratic bias, representative agent

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 are applied 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.[1]However, as Pollak points out, the premise that society has preferences that can be summarized by a social welfare function that 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 (Fisher & Griliches, 1995; Mackie & Schultze, 2002; Muellbauer, 1974; Nicholson, 1975; Pollak, 1980; Prais, 1959). The average can be unweighted (the so-called democratic index) or weighted, where the household indices are weighted according to that household'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 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.[2] First, households may face different prices. However, if the statistical system is such that the price data are 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.

A second issue is that households may be heterogeneous with respect to spending patterns. Statistical agencies typically report aggregate COLIs that are representative agent indices evaluated at economy-wide budget shares. Such indices are more representative of the consumption patterns of the higher income groups. Research has called for remedies either in terms of indices for subpopulations or a democratic aggregate index. But if these remedies are difficult to apply, then what is the bias caused by the use of representative agent indices? This is the problem studied in this paper. The precise problem is the following: if we accept the recommendation that the aggregate COLI should be an unweighted average (i.e., democratic) of individual/household indices, what would be the bias if the aggregate COLI was measured, instead, by computing the COLI for a representative agent, that is, the COLI that corresponds to average spending patterns. From previous work, we know that unless the expenditure function is of the Gorman polar form, a representative agent analysis is an invalid representation of the aggregate (Deaton & Muellbauer, 1980; Mackie & Schultze, 2002). The contribution here is to assess the direction and magnitude of bias for the important cases of the Tornqvist and Cobb–Douglas indices.

Commonly reported indices such as the Lowe index or the Laspeyres index[3] are linear in budget shares. Therefore, if they are aggregated in a plutocratic way, the resulting aggregate index is

[1] 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 & Pendakur, 2010).

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

[3] For most of the countries, reported CPI are Lowe price indices. It is a fixed basket index and the fixed basket usually corresponds to a period prior to the base period. The Laspeyres index is a special case of the Lowe index when the base period quantities constitute the fixed basket. Lowe and Laspeyres indices are linear in budget shares. Hence, the unweighted (democratic) average of the household indices equals the representative agent index with unweighted (democratic) average budget shares. Similarly, weighted (plutocratic) average of the household indices equals the representative agent index with weighted (plutocratic) average budget shares.

nothing but the representative agent index (with economy-wide budget shares). In these cases, the bias because of the use of a representative agent index is the same as the difference between a democratic representative agent index and a plutocratic representative agent index. This is the well-known plutocratic bias and has been extensively discussed (e.g., chapter 8 in Mackie & Schultze, 2002). However, when it comes to nonlinear indices, the difference identified above is only one component of the aggregation bias. There is a second component as well stemming from the curvature of the index. The paper shows that this component, in the case of Tornqvist and Cobb–Douglas indices, depends on the change in relative prices and the heterogeneity in budget shares.

In practice, statistical agencies typically report aggregate COLIs as representative plutocratic Lowe indices. The U.S. Bureau of Labor Statistics (BLS) carries over the practice of using the economy-wide budget shares in its construction of the Tornqvist index. 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 & 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. 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.

Our interest in the Tornqvist index comes from the fact that it is a superlative index (i.e., generated from an expenditure function of the flexible form) that is derived from a nonhomothetic translog expenditure function. The consistency with nonhomothetic 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- 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 aggregation bias[4] is composed of a curvature bias and the plutocratic bias. The curvature bias is always positive and depends on the heterogeneity in budget shares as well as the extent of change in relative prices. The bias is empirically evaluated for Indian and U.S. data for the representative agent indices of the Cobb–Douglas and the Tornqvist form. The bias in the Tornqvist index requires panel data, which are not commonly available. The paper proposes an upper bound to the bias that can be computed from repeated cross sections. The empirical exercises find the curvature bias to be small. Most of the aggregate bias stems from plutocratic bias, which can be eliminated by using democratic representative agent indices.

The remainder of the paper is organized as follows. Section 2 discusses the relevant literature. The decomposition of aggregation bias is explained in Section 3. The following section presents our findings of curvature bias and plutocratic bias for the Cobb–Douglas index using Indian and

[4] We have used the terms aggregation bias and representative agent bias synonymously in this paper.

U.S. data. Section 5 introduces the representative agent bias for the Tornqvist index using U.S. Quarterly Interview Survey data. In Section 6, we show the computation of the upper bound to the representative agent bias for the Tornqvist index. Section 7 concludes.

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 because of relative price changes.

Superlative indices are superior as they capture the substitution effect, which occurs due to the change in relative prices (Abraham et al., 1998; Boskin et al., 1998; Manser & McDonald, 1988). 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 nonhomothetic 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 (Fisher & Griliches, 1995; Mackie & Schultze, 2002; Muellbauer, 1974; Nicholson, 1975; Pollak, 1980; Prais, 1959).

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 & Schultze, 2002). Aggregation by way of an unweighted average—that is, a democratic index—reduces the bias that exists in a plutocratic representative index towards the consumption patterns of the better off. However, a democratic index requires the 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 (Cage et al., 2002; Crawford, 1994; Crawford & Smith, 2002; Del Río & Ruiz-Castillo, 2002; Garner et al., 1996, 2003; Hagemann, 1982; Idson & Miller, 1994; Kokoski, 1987; Livada, 1990; Michael, 1979;). 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 subgroups of the population like the elderly (Hobijn & Lagakos, 2003; Stewart, 2008) and for different demographic and income groups (Lyssiotou & 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 U.S. 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 unweighted or democratic average of these household COLIs. However, if statistical agencies followed the practice of using economy-wide budget shares, they would arrive at the COLI of the representative agent. How well does this approximate the aggregate COLI?

As mentioned earlier, our analysis considers the Tornqvist index. The U.S. 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 the 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 | COMPONENTS OF AGGREGATION BIAS

Consider a population of N households. We measure the change in the cost of living for each household by the Tornqvist index defined over M commodities.[5] For the jth household, let $s_i^{1,j}$ and $s_i^{0,j}$ be the budget shares for the ith commodity at period 1 and period 0, respectively. Define the average budget share as

$$s_i^j = \left(\frac{1}{2}\right)\left(s_i^{1,j} + s_i^{0,j}\right) \ \forall i = 1,2,\ldots,M \text{ and } j = 1,2,\ldots,N.$$

[5] The Tornqvist index is generated from a flexible and nonhomothetic translog expenditure function (Diewert, 1976). The expenditure function for the jth household is of the following form:$\ln C^j(u,P) = a_0^j + \sum_{i=1}^M a_i^j \ln P_i + (\frac{1}{2})\sum_{i=1}^M \sum_{k=1}^M a_{ik}^j \ln P_i \ln P_k + b_0^j \ln u^j + \sum_{i=1}^M b_i^j \ln P_i \ln u^j + (\frac{1}{2}) b_{00}(\ln u^j)^2$. The parameters satisfy the following restrictions:$a_{ik}^j = a_{ki}^j \ \forall i = 1,2,\ldots,M$, $k = 1,2,\ldots,M$ and $j = 1,2,..,N$; $\sum_{i=1}^M a_i^j = 1$; $\sum_{i=1}^M b_i^j = 0$; $\sum_{k=1}^M a_{ik}^j = 0 \ \forall i = 1,2,\ldots,M$ and $j = 1,2,..,N$.

Then the Tornqvist index for the jth household is

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

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

Without loss of generality, assume the Mth commodity to be the numeraire commodity.[6] We denote λ_i to be the ratio of the relative price of commodity i in period 1 to its relative price in period 0; that is

$$\frac{P_i^1}{P_M^1} / \frac{P_i^0}{P_M^0} = \lambda_{\mathrm{i}} \forall i = 1, 2, \ldots, M$$

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

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

The aggregate index for this population is the average of the household Tornqvist indices. While our interest is in the democratic unweighted average, the aggregation can also be weighted by the household's share in total expenditures (the plutocratic average). 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 the vector s denote a particular allocation of budget shares, $(s_1, s_2, \ldots, s_M)$. Let $\mathrm{B} = \{s : \sum_{m=1}^{M} s_m = 1\}$ denote the set of all possible allocations of budget shares. If $\boldsymbol{s}^j$ denotes the budget share allocation of the jth household, define the indicator function:

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

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

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

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

[6] As we shall see, this makes the bias expression interpretable in terms of relative prices.

where $\sum_{s \in B} d(s) = 1$ We call $d(s)$ as the democratic density.

Hence, the democratic aggregate COLI is defined by

$$A_d \equiv \sum_{s \in B} T(s) d(s) = E_d(T(s)), \tag{1}$$

where the expectations operator is indexed by d to remind us that the averaging is democratic. The corresponding democratic representative agent index is

$$R_d = T(E_d(s)) \tag{2}$$

For the plutocratic group COLI, we define the 'plutocratic' density as

$$p(s) = \sum_{j=1}^{N} \left(\frac{C^j}{\sum_{j=1}^{N} C^j} \right) V^j(s) \quad \forall s \in B$$

where C^j is the total expenditure made by the 'j'th household. Clearly $\sum_{s \in B} p(s) = 1$. The plutocratic group cost of living index is defined as

$$A_p = \sum_{s \in B} T(s) p(s) = E_p(T(s)), \tag{3}$$

where the expectations operator is indexed by p to remind us that the averaging is plutocratic. The corresponding plutocratic representative agent index is

$$R_p = T\left(E_p(s)\right). \tag{4}$$

The problem can now be clearly seen. The ideal aggregate index is A_d, but what the statistical agencies report is R_p. The bias, in percentage terms, in the plutocratic representative agent index is then defined as

$$g = \frac{A_d - R_p}{R_p} = \frac{E_d(T(s)) - T\left(E_p(s)\right)}{T\left(E_p(s)\right)}.$$

This can be decomposed into two terms as follows[7]:

$$g = \frac{E_d(T(s)) - T(E_d(s))}{T\left(E_p(s)\right)} + \frac{T(E_d(s)) - T\left(E_p(s)\right)}{T\left(E_p(s)\right)}. \tag{5}$$

The second term on the right-hand side of (5) arises because of the difference between a democratic and plutocratic weighting of commodity budget shares over households. Hence, this can be called the plutocratic bias. If households had identical shares of total economy-wide expenditure, the plutocratic weighting coincides with the democratic weighting and the second term disappears. The other case when it is negligible is when the relative price changes are the same for all households, rich or poor, which would happen if the change in prices does not differ across

[7] The expression of the bias reported in Equation (5) needs to be multiplied by 100 to get the percentage figures.

commodities. Other than these cases, the plutocratic bias is a source of aggregation bias for all price indices including the ones that are linear in budget shares (e.g., Lowe and Laspeyres). On the other hand, for the Tornqvist index, the following may be noted:

Proposition 1:$T(s)$ is convex in s.

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

Proposition 2: $E_d[T(s)] \geq T[E_d(\mathbf{s})]$.

This result shows that the first component of the bias is positive. As this is because of the convexity of the index, the first component can be called the curvature bias. Controlling for the plutocratic bias, the curvature bias leads the representative agent approximation to 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 & Stiglitz, 1970), increases the aggregate COLI. For linear price indices, the curvature bias vanishes and only the plutocratic bias remains.

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. Denoting the curvature bias as g_1, this can be rewritten as

$$g_1 = \frac{E_d\left[T\left(s\right)\right] - T\left[E_d\left(s\right)\right]}{T\left[E_d\left(s\right)\right]} \frac{T[E_d\left(s\right)]}{T\left[E_p\left(s\right)\right]}.$$

Proposition 3: The curvature bias can be approximated by the following:

$$g_1 \equiv \frac{E_d\left[T\left(s\right)\right] - T\left[E_d\left(s\right)\right]}{T\left[E_d\left(s\right)\right],} \frac{T[E_d\left(s\right)]}{T\left[E_p\left(s\right)\right]} \approx \left(\frac{1}{2}\right) \mathrm{var}_d \left[\sum\nolimits_{i=1}^{M-1} S_i \ln\lambda_i\right] \frac{T[E_d\left(s\right)]}{T\left[E_p\left(s\right)\right]}, \tag{6}$$

where var stands for variance and is subscripted by d to indicate that it is measured with respect to the democratic density.

For a proof of this result, see the Appendix. The expression in (6) is clearly nonnegative. The curvature 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 $\lambda_i = 1\ \forall i = 1, 2, \ldots, M-1$.[8] Computing the curvature bias requires panel data at the household level to obtain information about the base and current period shares.

The counterpart of Equation (6) for the Cobb–Douglas price index is

$$g_2 \approx \left(\frac{1}{2}\right) \mathrm{var}_d \left[\sum\nolimits_{i=1}^{M-1} \alpha_i \ln\lambda_i\right] \frac{C\left(E_d\left(\alpha\right)\right)}{C\left(E_p\left(\alpha\right)\right)}, \tag{7}$$

where α_i is the fixed budget share of ith commodity and $\frac{C(E_d(\alpha))}{C(E_p(\alpha))}$ is the ratio of the Cobb–Douglas indices evaluated at the democratic and plutocratic average budget shares. The bias in (7) can be estimated from cross-sectional data alone.[9]

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

[9] The Cobb–Douglas price index is also called the geometric Laspeyres index (Balk, 2009). It can be considered as a geometric version of the Lowe index with updated weights from a third period.

TABLE 1 Budget share of commodities (Indian data: Rural)

Commodity	Mean	Standard deviation	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 nonfood	0.22	0.12	55

Note: Authors' calculation from National Sample Survey (2004–2005) data. The mean budget shares as well as measures of dispersion are evaluated by democratic density. CV stands for coefficient of variation.

The curvature bias component closely resembles the difference between a Carli and Jevons index derived by Diewert (2004). In the aggregation of price ratios, the Carli index is the arithmetic average while the Jevons index is the geometric average of the price ratio/price relatives (Diewert, 2004). Indeed, as pointed out by a referee, the curvature bias vanishes if the aggregate COLI was a geometric average of individual indices rather than an arithmetic average.

4 | REPRESENTATIVE AGENT BIAS IN THE COBB–DOUGLAS INDEX

We begin by presenting the bias estimates for the Cobb–Douglas index (i.e., Equation 7). For this purpose, we use cross-sectional data from India and the United States.

4.1 | India

The nationally representative consumer expenditure survey of 2004–2005 is used, which samples about 120,000 households across rural and urban India. Following Almås and Kjelsrud (2017), we classify all expenditures 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 measures of dispersion—both evaluated by the democratic density. Notice that the coefficient of variation is more than 100% or close to 100% for a few of the commodities. Such heterogeneity is not peculiar to the Indian dataset.[10]

[10] 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 United States 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. (1996) construct an

TABLE 2 Budget share of commodities (Indian data: Urban)

Commodity	Mean	Standard deviation	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 nonfood	0.33	0.15	45

Note: Authors' calculation from National Sample Survey (2004–2005) data. The mean budget shares as well as measures of dispersion are evaluated by democratic density. CV stands for the coefficient of variation.

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 nonfood, which is considered as a numeraire good) for all categories between 2004–2005 and 2011–2012.[11] Scenarios 2 and 3 are hypothetical. In Scenario 2, we suppose the percentage price changes are highest for the commodities consumed largely by the poor.[12] 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 nonfood). In this case, we would expect the democratic representative agent index to rise more than the counterpart plutocratic index and hence the plutocratic bias to be positive. 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%. The plutocratic bias is expected to be negative in this case. All these three scenarios can be compared with a 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.

The bottom three rows of Tables 3 and 4 display the bias calculations. The curvature bias is, as expected positive. However, it turns out to be very small and is comparable across the three scenarios. The plutocratic bias is larger by several orders of magnitude and expectedly varies substantially across the scenarios. For this reason, the overall bias is larger—ranging between −1% and 1.5% for the other cases because of plutocratic bias.

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 United Kingdom 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.

[11] 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.

[12] These are the food categories of "cereals and cereal substitutes" and "pulse and pulse products."

TABLE 3 Change in prices and representative agent bias for the Cobb–Douglas index (Indian data: Rural)

	Change in prices (%)		
Commodity	**Scenario 1**	**Scenario 2**	**Scenario 3**
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 nonfood	64	20	80
Curvature bias (%)	0.06	0.07	0.07
Plutocratic bias (%)	0.47	1.22	−1.21
Overall bias (%) = curvature bias (%)+ plutocratic bias (%)	0.53	1.29	−1.14

Note: In Scenario 1, we consider the observed change in prices for all categories between 2011−2012 and 2004−2005. The changes shown in the table are percentage changes in prices. In Scenario 2, we consider two different rates of change in 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 all other commodity groups increase at a rate of 80%. .

4.2 | United States

In the United States, the consumer expenditure survey is conducted by the BLS. We use the Quarterly Interview Survey for the period 2015–2018. It is a rotating panel where 25% of the existing consumer units are rotated out every quarter, and so every unit reports quarterly expenditures for a year.[13]

Table 5 reports the average budget share and its variation for eight consumption categories. The numbers are displayed for the second quarter of 2015, but the magnitudes are similar across other years and quarters. The budget shares exhibit substantial heterogeneity much like the Indian case.

The computation of the aggregation bias disaggregates the eight consumption categories reported in Table 5 to 41 categories. Like in the Indian case, we consider three scenarios of price change—the observed price change between the second quarter of 2015 and the second quarter of

[13] Any household/consumer unit is asked to report their expenditure for the last 3 months. For example, consider the second quarter for any particular year. If a consumer unit is interviewed in May, it reports expenditures for February, March, and April. The expenditure incurred in February and March is the last quarter expenditure and expenditure in April is the current quarter expenditure. Therefore, those who are interviewed in April do not report any expenditure for the current (i.e., second) quarter. On the other hand, the consumer units interviewed in June report their last quarter expenditure for the month of March and current quarter expenditure for the month of April and May. Therefore, in order to calculate the total expenditure reported in any quarter by each consumer unit, the expenditure made in the last and current quarter needs to be added. It is also called the “collection period” expenditure.

TABLE 4 Change in prices and representative agent bias for the Cobb–Douglas index (Indian data: Urban)

	Change in prices (%)		
Commodity	**Scenario 1**	**Scenario 2**	**Scenario 3**
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 nonfood	47	20	80
Curvature bias	0.05	0.06	0.06
Plutocratic Bias (%)	1.33	1.18	−1.17
Overall bias (in %) = curvature bias (%)+ plutocratic bias (%)	1.38	1.24	−1.11

Note: In Scenario 1, we consider the observed change in prices for all categories between 2011–2012 and 2004–2005. The changes shown in the table are percentage changes in prices. In Scenario 2, we consider two different rates of change in 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 all other commodity groups increase at a rate of 80%.

2018 and two hypothetical scenarios.[14] The observed price changes for the 41 categories between the second quarter of 2015 and the second quarter of 2018 are the corresponding price indices of these categories computed from the price data provided by the BLS. The first column of Table 6 reports the observed price change when aggregated into eight categories. The scenario in the second column assumes that all commodities within the food and beverages, shelter, and utility categories experience a price change of 8%, while it is 2% for all other commodities. In Scenario 3, the price change scenarios are reversed. The last three rows report the bias calculations.

As expected, the plutocratic bias reverses in sign between the second and the third scenarios. Like the Indian case, the curvature bias is small and dominated by the plutocratic bias. These results are similar when they are computed for price changes between 2015 and 2018 for the other quarters.

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. And that is the case with the

[14] The results from the first scenario are similar for observed price changes between 2015 and 2018 for other quarters. It also makes no difference whether the prices are seasonally adjusted or not.

TABLE 5 Budget share of commodities (US data: second quarter, 2015)

Commodity	Mean	Standard deviation	CV (%)
Food and beverages	0.19	0.1	53
Shelter	0.23	0.15	68
Utilities	0.14	0.09	64
Apparel	0.02	0.03	150
Transport and vehicles	0.13	0.14	108
Health and health-related services	0.08	0.1	125
Entertainment	0.04	0.05	125
Other miscellaneous expenditure	0.17	0.13	76

Note: Authors' calculation from the second quarter of the Quarterly Interview Survey, 2015. The figures for the other quarters are almost the same. Each of these categories has many subcategories. The budget shares of these subcategories have been directly used to compute representative agent bias. There are 41 categories in the disaggregated data that we have used for computing representative agent bias. In this table, we show the summary figures for the eight aggregated categories constructed from the disaggregated ones. The mean budget shares as well as the measures of dispersion are evaluated by democratic density. CV stands for the coefficient of variation.

TABLE 6 Change in prices and representative agent bias for the Cobb–Douglas index (US data: Quarterly Interview Survey)

	Changes in prices (%)		
Commodity	Scenario 1	Scenario 2	Scenario 3
Food and beverages	3	8	2
Shelter	9	8	2
Utilities	1	8	2
Apparel	1	2	8
Transport and vehicles	4	2	8
Health and health-related services	7	2	8
Entertainment	4	2	8
Other miscellaneous expenditure	7	2	8
Curvature bias (%)	0.02	0.005	0.005
Plutocratic bias (%)	0.51	0.38	−0.38
Overall bias (%) = Curvature bias (%)+ Plutocratic Bias (%)	0.53	0.385	−0.375

Note: The price changes have been computed from the price indices of individual commodities/categories available from the Bureau of Labor Statistics. There are 41 categories used in the computation of the representative agent bias. But the above table aggregates them into eight categories by taking the geometric average of the change in prices of the individual subcategories under each of these eight categories. Scenario 1 corresponds to the actual change in prices between the second quarter of 2018 and the second quarter of 2015. In Scenario 2, we consider two different rates of change in prices. The prices of the necessities (food and beverages, shelter, and utilities) are assumed to increase at a rate of 8%. The prices of other categories are assumed to increase at a rate of 2%. Scenario 3 is the exact opposite of Scenario 2, where the prices of the necessities increase at a rate of 2% and the prices of other commodity groups increase at a rate of 8%. All the reported figures are in percentages.

household surveys in India which are cross sectional. Panels can be constructed from the U.S. Quarterly Interview Survey where households are surveyed for four quarters before they are rotated out. We construct four panels for the adjacent years 2015–2016. The first panel uses the first quarter of 2015 (Q1, 2015) as the base period and the fourth quarter of 2015 (Q4, 2015) as the

TABLE 7 Change in prices and representative agent bias for the Tornqvist index (US data: Quarterly Interview Survey)

Commodity	Changes in prices (%)			
	Panel 1	Panel 2	Panel 3	Panel 4
Food and beverages	1	1	1	0
Shelter	2	2	3	2
Utilities	−5	−5	−5	−2
Apparel	0	0	0	0
Transport and vehicles	−1	−1	−3	−1
Health and health-related services	1	2	2	3
Entertainment	1	1	1	2
Other miscellaneous expenditure	1	1	2	1
Curvature bias (%)	0.002	0.003	0.006	0.003
Plutocratic bias (%)	−0.072	0.009	−0.058	−0.027
Overall bias (%) = curvature bias (%)+ plutocratic bias (%)	−0.07	0.012	−0.052	−0.024

Note:-The price changes have been computed from the price indices of individual commodities/categories available from the Bureau of Labor Statistics. There are 41 categories used in the computation of the representative agent bias. But the above table aggregates them into eight categories by taking the geometric average of the change in prices of the individual subcategories under each of these eight categories. The price changes for any panel are the change in prices between the quarters that constitute the panel. Any negative price change implies a decline in the price of that commodity/category. Panel 1 is constituted of those households who are interviewed both in the first and fourth quarters of 2015. Similarly, in Panel 2 we have households who are interviewed both in the second quarter of 2015 and the first quarter of 2016. Panel 3 is formed of the third quarter of 2015 and second quarter of 2016. Panel 4 is formed of the fourth quarter of 2015 and third quarter of 2016. All the reported figures are in percentages.

current period. Continuing this way, we have three other panels (Q2, 2015–Q1, 2016; Q3, 2015–Q2, 2016; Q4, 2015–Q3, 2016). Panels extending for more than a year are not possible.

For the eight major consumption categories, Table 7 reports the price change between the base period and the current period (computed as the geometric average of the change in prices for commodities within the category). The last three rows of Table 7 display the components of aggregation bias and the overall bias. These bias calculations are based not on the eight category classification but on a much more detailed disaggregation of 41 consumption categories.

Given the limited change in relative prices in this panel, the curvature bias can be expected to be small. And that is the case. As in the earlier results, the plutocratic bias is much greater, often by a factor of 10 or more. Nonetheless, the overall bias is still very low.

6 | AN UPPER BOUND TO THE BIAS IN THE TORNQVIST INDEX

As noted earlier, panel data on consumption expenditure are not commonly available. In the United States, for instance, the Quarterly Interview Survey provides a comprehensive dataset on the spending habits of U.S. 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 or the Health and Retirement Survey, have abundant information on income or wealth, but no information whatsoever on consumption. In the United

Kingdom, the Family Expenditure Survey 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, typically lack consumption data.

While the absence of panel data constrains the computation of individual and aggregate Tornqvist indices, it does not constrain the calculation of representative agent Tornqvist indices, whether democratic or plutocratic. The representative agent indices can be calculated from repeated cross sections because they require only averages. The absence of panel data is a problem for the curvature bias alone. However, from cross-section data, we can compute an upper bound to the curvature bias. This is what we do in this section.

From (6), the curvature bias for the Tornqvist index can be expressed as

$$g_1 \approx \left[\left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \text{var}_d\,(s_i)\,(\ln\lambda_i)^2 + \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} \text{cov}_d(s_i s_k)(\ln\lambda_i)(\ln\lambda_k) \right] \frac{T[E_d(s)]}{T[E_p(s)]}, \tag{8}$$

where the budget shares are averages over the base (period 0) and current period (period 1). Thus, both the variance and the covariance terms above require household budget share data for a base and a current period. For the variance terms, we show in the appendix that

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

where the right-hand side can now be computed by cross-sectional data for the base and current periods. Using the Cauchy–Schwartz inequality, the covariance terms in Equation (8) can also receive an upper bound. We show in the Appendix that

$$\begin{aligned} &\sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} cov_d(s_i s_k)(\ln\lambda_i)(\ln\lambda_k) \\ &\quad \leq \left(\frac{1}{4}\right) \sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} \left[cov_d\left(s_i^0, s_k^0\right) + cov_d\left(s_i^1, s_k^1\right)\right] (\ln\lambda_i)(\ln\lambda_k) \\ &\quad + \left(\frac{1}{4}\right) \sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} \left[\sqrt{var_d\left(s_i^1\right) var_d\left(s_k^0\right)} + \sqrt{var_d\left(s_i^0\right) var_d\left(s_k^1\right)}\right] (\ln\lambda_i)(\ln\lambda_k) \end{aligned}$$

Therefore, an upper bound to the curvature bias is derived as

$$\begin{aligned} g_1 &\approx [\left(\tfrac{1}{2}\right) \textstyle\sum_{i=1}^{M-1} \text{var}_d\,(s_i)(\ln\lambda_i) + \left(\tfrac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} \text{cov}_d\,(s_i, s_k)](\ln\lambda_i)(\ln\lambda_k)\frac{T[E_d(s)]}{T[E_p(s)]} \\ &\leq \left(\tfrac{1}{8}\right) \textstyle\sum_{i=1}^{M-1} \left[\text{var}_d\left(s_i^1\right) + \text{var}_d\left(s_i^0\right) + 2\sqrt{\text{var}_d\left(s_i^1\right)\text{var}_d\left(s_i^0\right)}\right](\ln\lambda_i)^2 \frac{T[E_d(s)]}{T[E_p(s)]} \\ &\quad + \left(\tfrac{1}{8}\right) \textstyle\sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} \left[\text{cov}_d\left(s_i^0, s_k^0\right) + \text{cov}_d\left(s_i^1, s_k^1\right)\right](\ln\lambda_i)(\ln\lambda_k)\frac{T[E_d(s)]}{T[E_p(s)]} \\ &\quad + \left(\tfrac{1}{8}\right) \textstyle\sum_{i=1}^{M-1} \sum_{k=1, i\neq k}^{M-1} \left[\sqrt{\text{var}_d\left(s_i^1\right)\text{var}_d\left(s_k^0\right)} + \sqrt{\text{var}_d\left(s_i^0\right)\text{var}_d\left(s_k^1\right)}\right](\ln\lambda_i)(\ln\lambda_k)\frac{T[E_d(s)]}{T[E_p(s)]}. \end{aligned} \tag{9}$$

TABLE 8 Upper bound to the representative agent bias for the Tornqvist index: India

	upper bound on curvature bias (%)	Plutocratic bias (%)	Upper bound on the overall bias (%)
Rural	0.08	0.53	0.61
Urban	0.14	1.3	1.44

Note: The calculations are based on the National Sample Survey (2004–2005 and 2011–2012). These upper bounds correspond to the observed price changes between 2011–2012 and 2004–2005, that is,. price changes under Scenario 1 in Tables 3 and 4.

TABLE 9 Upper bound to the representative agent bias for the Tornqvist index: United States

	Upper bound on curvature bias (%)	Plutocratic bias (%)	Upper bound on the overall bias (%)
Q2, 2015–Q2, 2018	0.07	0.56	0.63
Panel 1	0.009	−0.072	−0.063
Panel 2	0.011	0.009	0.02
Panel 3	0.007	−0.058	−0.051
Panel 4	0.005	−0.027	−0.022

Note:-The calculations are based on U.S. Quarterly Interview Survey. The first row in this table shows the upper bounds corresponding to the observed changes in prices between the second quarter of 2018 and the second quarter of 2015, that is, Scenario 1 in Table 6. The four panels are the ones mentioned in Table 7. The price changes for any panel are the observed changes in prices between the quarters that constitute the panel. Between the second quarter of 2018 and the second quarter of 2015, we cannot directly compute the exact Tornqvist representative agent bias and just compute the upper bound to the exact bias. For all four panels, we can compute the exact Tornqvist representative agent bias (shown in Table 7) as well as the corresponding upper bounds.

The right-hand side of expression (9) can be solely computed from cross-sectional data in the base and current periods. Thus, when restricted to cross-sectional data, the upper bound to the aggregate bias would then be the sum of the exact plutocratic bias and the upper bound to the curvature bias.

Table 8 displays the estimates of the upper bound to the representative agent bias in the Tornqvist index for the Indian data. These are computed for the observed price changes between 2004–2005 and 2011–2012, that is, Scenario 1 of Tables 3 and 4. The overall bias is less than 1.5% of which the contribution of the curvature bias is no more than one-tenth.

Table 9 displays the corresponding estimates for the United States. The first row computes the upper bound to the bias for the change in prices between the second quarter of 2015 and the second quarter of 2018 (called Scenario 1 in Table 6).[15] For this period, panel data were not available and so the upper bound calculations are of value. The other rows correspond to the panels constructed in the earlier section. These panels were used to calculate the exact bias in the Tornqvist index, and the upper bound estimates here provide a useful comparison. The estimates confirm the general pattern: that the overall bias is small even over extended periods and that most of it come from plutocratic bias.

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

[15] The estimates are similar for other quarters.

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 the bias has two components: the curvature bias and the plutocratic bias. The latter is well recognized in the literature but not the former. For an important and widely used superlative index like the Tornqvist, the nature of the curvature 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 paper also estimates the extent of this bias for Indian and the U.S. data. An upper bound to the bias for the Tornqvist index can be found from cross-sectional data alone which is otherwise insufficient to estimate the exact aggregation bias corresponding to the Tornqvist index. The empirical exercises show that the magnitude of the curvature bias is small. The plutocratic bias is the dominant source usually accounting for 90% or more of the total bias.

For India, when we consider the observed change in prices between 2004–2005 and 2011–2012, the overall Cobb–Douglas representative agent bias (expressed in percentage terms) turns out to be 0.53% and 1.38% for the rural and urban sample, respectively. The overall bias is 0.53% in the U.S. data between the second quarter of 2015 and the second quarter of 2018. The contribution of the curvature bias in the total bias never exceeds 10%, and the rest is explained by the plutocratic bias. Similarly, plutocratic bias also turns out to be the major component of the Tornqvist representative agent bias computed from the U.S. quarterly panel data and often greater by a factor of 10 or more relative to the curvature bias.

The difference in the relative size of these two sources of bias is because of a couple of factors. While both the curvature bias and the plutocratic bias are evaluated for the same change in relative prices, there is a difference in how they enter the respective expressions. In curvature bias, relative price changes appear in logarithmic terms while they appear in absolute terms for plutocratic bias. Logarithmic transformation reduces the magnitude. Second, while plutocratic bias depends on the difference in plutocratic and democratic average budget shares, the curvature bias depends on the variances and covariances of budget shares. As budget shares lie between zero and one, these variances and covariances terms turn out to be small.

It implies that, in practice, much of the aggregate bias can be removed by using democratic budget shares and reporting the democratic representative agent index.

ORCID

Sutirtha Bandyopadhyay https://orcid.org/0000-0002-5049-127X

REFERENCES

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

Almås, I., & Kjelsrud, A. (2017). Rags and riches: Relative prices, non-homothetic preferences, and inequality in India. *World Development*, *97*, 102–121.

Balk, B. M. (2009). Lowe and Cobb-Douglas CPIs and their substitution bias. Paper presented at Proceedings of the Eleventh Meeting of the International Working Group on Price Indices (Neuchâtel, Switzerland). www.ottawagroup.org

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

Cage, R. A., Garner, T., & Ruiz-Castillo, J. (2002). *Constructing household specific consumer price indexes: An analysis of different techniques and methods* (BLS Working Paper No. 354). U.S. Bureau of Labor Statistics.

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

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

Crawford, I., & Smith, Z. (2002). Distributional aspects of inflation. *(Commentary no. 90)*. Institute for Fiscal Studies.

Crossley, T. F., & Pendakur, K. (2010). The common-scaling social cost-of-living index. *Journal of Business & Economic Statistics*, *28*(4), 523–38.

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

Del Río, C., & Ruiz-Castillo, J. (2002). Demographic trends and living standards: The case of Spain during the 1980s. *Revista de Economía Aplicada*, *X*, 217–234.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, *4*(2), 115–245.

Diewert, W. E. (1978). Superlative index numbers and consistency in aggregation. *Econometrica*, *46*(4), 883–900.

Diewert, W. E. (2004). Basic index number theory. In Paul A. Arm-Knecht (Ed.), *Producer price index manual: Theory and practice.*. International Labour Office, International Monetary Fund, Eurostat, United Nations, and The World Bank, and Organisation for Economic Co-operation and Development.

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

Fisher, F. M., & Griliches, Z. (1995). Aggregate price indices, new goods, and generics. *Quarterly Journal of Economics*, *110*(1), 229–244.

Garner, T. I., D. S. Johnson, & . F. Kokoski. (1996). An experimental consumer price index for the poor.' *Monthly Labor Review*, *119*:(9), 32–42

Garner, T. I., Ruiz-Castillo, J., & 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), 22–48.

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

Hobijn, B., & Lagakos, D. (2003). Social security and the consumer price index for the elderly. *Current Issues in Economics and Finance*, *9*(5), 1–7.

Idson, T., & 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, Department of Economics.

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

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

Livada, A. (1990). The distribution of household inflation rates: The Greek experience. *Bulletin of Economic Research*, *42*(3), 175–196.

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

Mackie, C., & Schultze, C. L. (2002). *At what price? Conceptualizing and measuring cost-of-living indexes*. National Academy Press.

Manser, M. E., & McDonald, R. J. (1988). An analysis of substitution bias in measuring inflation, 1959–85. *Econometrica*, *56*(4), 909–930.

Michael, R. T. (1979). Variation across households in the rate of inflation. *Journal of Money, Credit and Banking*, *11*(1), 32–46.

Muellbauer, J. . *The political economy of price indices* (Birkbeck Discussion Paper no. 22). Birkbeck, University of London.

Nicholson, J. L. (1975). Whose cost of living? *Journal of the Royal Statistical Society. Series A (General)*, *138*(4), 540–542.

Pollak, R. A. (1980). Group cost-of-living indexes. *American Economic Review*, *70*(2), 273–278.
Pollak, R. A. (1981). The social cost of living index. *Journal of Public Economics*, *15*(3), 311–336.
Porto, G. G. (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics*, *70*(1), 140–60.
Prais, S. J. (1959). Whose cost of living? *Review of Economic Studies*, *26*(2), 126–134.
Prud'homme, M., Sanga, D., & Yu, K. (2005). A computer software price index using scanner data. *Canadian Journal of Economics/Revue Canadienne d'économique*, *38*(3), 999–1017.
Rothschild, M., & Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, *2*(3), 225–43.
Stewart, K. J. (2008). The experimental consumer price index for elderly Americans (CPI-E): 1982–2007. *Monthly Labor Review*, *131*(4), 19–24

How to cite this article: Bandyopadhyay S, Ramaswami B. The representative agent bias in cost of living indices. *Bull Econ Res* 2021;1–24. https://doi.org/10.1111/boer.12286

APPENDIX

Proof of Proposition 1

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

Hence the Hessian matrix can be written as

$$H = T\left(D.D^t\right),$$

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

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

Proof of Proposition 3

Considering the second-order Taylor's series expansion of $T(s)$ around $E_d(s)$, we obtain

$$\begin{aligned} T\ (s) = T\,[E_d\,(s)] + \sum_{i=1}^{M-1} (s_i - E_d\,(s_i))(\frac{\partial T}{\partial s_i}) + \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} (s_i - E_d\,(s_i))^2 \left(\frac{\partial^2 T}{\partial s_i^2}\right) \\ + \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1; i \neq k}^{M-1} (s_i - E_d\,(s_i))\,(s_k - E_d\,(s_k)) \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right) + R_2. \end{aligned} \tag{A1}$$

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

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

It can be shown that $R_2(E_d(s), h)$ is $o(||h||^2)$, that is, $\frac{R_2(E_d(s),h)}{||h||^2}$ tends to zero as 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

$$E_d\,[T\,(s)] - T\,[E_d\,(s)] \approx \left(\frac{1}{2}\right) \sum_{i\,=\,1}^{M-1} E_d[s_i - E_d\,(s_i)]^2 \left(\frac{\partial^2 T}{\partial s_i^2}\right)$$
$$+\left(\frac{1}{2}\right) \sum_{i\,=\,1}^{M-1} \sum_{k\,=\,1, i\neq k}^{M-1} E_d\,[(s_i - E_d\,(s_i))\,(s_k - E_d\,(s_k))] \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right) \tag{A2}$$
$$= \left(\frac{1}{2}\right) \left[\sum_{i\,=\,1}^{M-1} \mathrm{var}_d\,(s_i) \left(\frac{\partial^2 T}{\partial s_i^2}\right) + \sum_{i\,=\,1}^{M-1} \sum_{k\,=\,1, i\neq k}^{M-1} \mathrm{cov}_d\,(s_i, s_k) \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)\right].$$

Dividing both sides of Equation (A2) by $T[E_d(s)]$, we get

$$\frac{E_d\,[T\,(s)] - T\,[E_d\,(s)]}{T\,[E_d\,(s)]} \approx \left(\frac{1}{2}\right) \sum_{i\,=\,1}^{M-1} \mathrm{var}_d(s_i) \frac{\left(\frac{\partial^2 T}{\partial s_i^2}\right)}{T\,[E_d\,(s)]}$$
$$+ \left(\frac{1}{2}\right) \sum_{i\,=\,1}^{M-1} \sum_{k\,=\,1, i\neq k}^{M-1} \mathrm{cov}_d(s_i, s_k) \frac{\left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)}{T\,[E_d\,(s)]}. \tag{A3}$$

Now,

$$\frac{\left(\frac{\partial^2 T}{\partial s_i^2}\right)}{T[E_d(s)]} = (\ln \lambda_i)^2 \forall\ i\ =\ 1, 2, \ldots, M-1$$

and

$$\frac{\frac{\partial^2 T}{\partial s_i \partial s_k}}{T[E_d(s)]} = (\ln\lambda_i)(\ln\lambda_k) \forall\ i\ =\ 1, 2, \ldots, M-1;\ k\ =\ 1, 2, \ldots, M-1;\ i \neq \mathrm{k}.$$

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

$$\frac{E_d\,[T\,(s)] - T\,[E_d\,(s)]}{T\,[E_d\,(s)]} \approx \left(\frac{1}{2}\right) \sum_{i\,=\,1}^{M-1} \mathrm{var}_d(s_i)(\ln \lambda_i)^2$$
$$+ \left(\frac{1}{2}\right) \sum_{i\,=\,1}^{M-1} \sum_{k\,=\,1, i\neq k}^{M-1} \mathrm{cov}_d(s_i, s_k)(\ln \lambda_i)(\ln \lambda_k)$$
$$= \left(\frac{1}{2}\right) \mathrm{var}_d \left[\sum_{i\,=\,1}^{M-1} S_i \ln \lambda_i\right].$$

Therefore, the curvature bias is characterized by

$$g_1 \approx \left(\frac{1}{2}\right) \mathrm{var}_d \left[\sum_{i\,=\,1}^{M-1} S_i \ln \lambda_i\right] \frac{T[E_d\,(s)]}{T\,[E_p\,(s)]}. \tag{A4}$$

The expression (A4) is the same as Equation (6), 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

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

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

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

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

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

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

Now,

$$|\left(\frac{\partial}{\partial s_1}+\frac{\partial}{\partial s_2}+\cdots+\frac{\partial}{\partial s_{M-1}}\right)^3 T(t)| = |\sum_{i=1}^{M-1}\frac{\partial^3 T(t)}{\partial s_i^3}+3\sum_{i=1}^{M-1}\sum_{k=1;i\neq k}^{M-1}\frac{\partial^3 T(t)}{\partial s_i^2\partial s_k}$$
$$+\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}\sum_{l=1;i\neq k\neq l}^{M-1}\frac{\partial^3 T(t)}{\partial s_i\partial s_k\partial s_l}|.$$

Because we are considering the absolute value of the derivative, $(\frac{\partial}{\partial s_1}+\frac{\partial}{\partial s_2}+\cdots+\frac{\partial}{\partial s_{M-1}})^3 T(t)$, 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.

Derivation of the Upper Bound

The curvature bias for the Tornqvist index is expressed as

$$g_1 \approx [\left(\frac{1}{2}\right)\sum_{i=1}^{M-1}\mathrm{var}_d(s_i)(\ln\lambda_i)^2$$
$$+\left(\frac{1}{2}\right)\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}\mathrm{cov}_d(s_i s_k)(\ln\lambda_i)(\ln\lambda_k)]\frac{T[E_d(s)]}{T\left[E_p(s)\right]}$$
$$s_i=\left(\frac{1}{2}\right)\left(s_i^1+s_i^0\right);\quad s_k=\left(\frac{1}{2}\right)\left(s_k^1+s_k^0\right). \qquad \text{(A5)}$$

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. As the computation of the ratio $\frac{T[E_d(s)]}{T[E_p(s)]}$ does not require a panel, we only focus on the terms inside the square bracket of expression (A5) for the construction of the upper bound. Suppose we split up the expression for curvature bias (only the terms inside the square bracket) into two parts. The first part of the bias is

$$B1 = \left(\frac{1}{2}\right)\sum_{i=1}^{M-1} \mathrm{var}_d\,(s_i)(\ln\lambda_i)^2.$$

Now,

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

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

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

Now,

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

where R_i^2 is the squared correlation coefficient between s_i^1 and s_i^0 $\forall i = 1, 2, \ldots, M-1$. Replacing Replacing $\frac{cov_d(s_i^1, s_i^0)}{\sqrt{var_d(s_i^1)var_d(s_i^0)}}$ by $(R_i^2)^{\left(\frac{1}{2}\right)}$, we can write down the variance as

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

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:

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

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

$$B1 = \left(\frac{1}{2}\right)\sum_{i=1}^{M-1} \mathrm{var}_d\,(s_i)\,(\ln\lambda_i)^2$$

$$\leq \left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\left[\mathrm{var}_d\,\left(s_i^1\right)+\mathrm{var}_d\,\left(s_i^0\right)+2\sqrt{\mathrm{var}_d\,\left(s_i^1\right)\mathrm{var}_d\,\left(s_i^0\right)}\right](\ln\lambda_i)^2 \quad \text{(A6)}$$

Now we focus on the second term of the bias expression, which we can write as

$$B2 = \left(\frac{1}{2}\right)\sum_{i=1}^{M-1}\sum_{k=1, i\neq k}^{M-1} \mathrm{cov}_d(s_i, s_k)(\ln\lambda_i)(\ln\lambda_k).$$

Any covariance term $\mathrm{cov}_d(s_{i,}s_k)$ in the expression $B2$ can be rewritten in the following way:

$$\mathrm{cov}_d\;(s_i, s_k) = \mathrm{cov}_d\;\left(\left(s_i^1+s_i^0\right)/2, (s_k^1+s_k^0)/2\right)$$
$$= (1/4)\left[\mathrm{cov}_d\left(s_i^0, s_k^0\right)+\mathrm{cov}_d\left(s_i^1, s_k^0\right)+\mathrm{cov}_d\left(s_i^0, s_k^1\right)+\mathrm{cov}_d\left(s_i^1, s_k^1\right)\right].$$

The first and the fourth terms inside the square bracket can be readily computed from the cross-sectional data. By applying the Cauchy–Schwartz inequality, we can generate upper bounds on the second and third terms, that is, $cov_d(s_i^1, s_k^0)$ and $\mathrm{cov}_d(s_i^0, s_k^1)$. These terms can be bounded above as

$cov_d(s_i^1, s_k^0) \leq \sqrt{\mathrm{var}_d(s_i^1)\mathrm{var}_d(s_k^0)}$ and $\mathrm{cov}_d(s_i^0, s_k^1) \leq \sqrt{\mathrm{var}_d(s_i^0)\mathrm{var}_d(s_k^1)}$.
Therefore, the upper bound on the entire covariance term $B2$ can be written as

$$\left(\frac{1}{2}\right)\sum_{i=1}^{M-1}\sum_{k=1, i\neq k}^{M-1}\mathrm{cov}_d(s_i s_k)(\ln\lambda_k)$$

$$\leq \left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\sum_{k=1, i\neq k}^{M-1}\left[\mathrm{cov}_d\left(s_i^0, s_k^0\right)+\mathrm{cov}_d\left(s_i^1, s_k^1\right)\right](\ln\lambda_i)(\ln\lambda_k) \quad \text{(A7)}$$

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

Combining (A6) and (A7), the upper bound on the entire curvature bias term is written as

$$g_1 \approx \left[\left(\frac{1}{2}\right)\sum_{i=1}^{M-1}\mathrm{var}_d\,(s_i)+\left(\frac{1}{2}\right)\sum_{i=1}^{M-1}\sum_{k=1, i\neq k}^{M-1}\mathrm{cov}_d\,(s_i, s_k)\right](\ln\lambda_i)(\ln\lambda_k)\frac{T[E_d\,(s)]}{T\,[E_p\,(s)]}$$

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

$$+\left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}\left[\mathrm{cov}_d\left(s_i^0,s_k^0\right)+\mathrm{cov}_d\left(s_i^1,s_k^1\right)\right](\ln\lambda_i)(\ln\lambda_k)\frac{T[E_d(s)]}{T\left[E_p(s)\right]} \tag{A8}$$

$$+\left(\frac{1}{8}\right)\sum_{i=1}^{M-1}\sum_{k=1,i\neq k}^{M-1}\left[\sqrt{\mathrm{var}_d\left(s_i^1\right)\mathrm{var}_d\left(s_k^0\right)}+\sqrt{\mathrm{var}_d\left(s_i^0\right)\mathrm{var}_d\left(s_k^1\right)}\right](\ln\lambda_i)(\ln\lambda_k)\frac{T[E_d(s)]}{T\left[E_p(s)\right]}.$$

The upper bound on the curvature bias, that is the right-hand side of expression (A8) is exactly the same as shown in Equation (9) in the main text. The upper bound can solely be computed from the cross-sectional data on base period and current period budget shares.