



ASHOKA
UNIVERSITY

Ashoka University Economics
Discussion Paper 49

Representative Agent Bias in Cost of Living Indices

January 2021

Sutirtha Bandyopadhyay, Indian Institute of Management, Indore
Bharat Ramaswami, Ashoka University

<https://ashoka.edu.in/economics-discussionpapers>

The Representative Agent Bias in Cost of Living Indices

Sutirtha Bandyopadhyay¹ and Bharat Ramaswami²

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 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: Cost of Living Index, Aggregation, Heterogeneity, Representative Agent, Curvature Bias, Plutocratic Bias

¹Assistant Professor, Indian Institute of Management Indore ; email: sutirtha.bandyopadhyay@gmail.com

²Professor, Ashoka University, Delhi ; 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 expositied 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 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

³Another 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).

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.⁴First, households may face different 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.

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 sub-populations 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, i.e., 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 (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.

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

Commonly reported indices such as the Lowe index or the Laspeyres index⁵ are linear in budget shares. Therefore, if they are aggregated in a plutocratic way, the resulting aggregate index is 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 and Schultze, 2002). However, when it comes to non-linear 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 cost of living indices as representative plutocratic Lowe 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. 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. We are not aware of any other country reporting a superlative COLI. But if they plan to go in that direction, then they too

⁵For 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. 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.

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 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 aggregation bias⁶ 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 US data for the representative agent indices of the Cobb-Douglas and the Tornqvist form. The bias in the Tornqvist index requires panel data which is 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

⁶ We have used the terms aggregation bias and representative agent bias synonymously in this paper.

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 Cobb-Douglas index using Indian and US data. Section 5 introduces the representative agent bias for Tornqvist index using US quarterly interview survey data. In section 6, we show the computation of the upper bound to the representative agent bias for 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 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 representative 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 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 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. Components of Aggregation Bias

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.⁷For 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

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

Then the Tornqvist index for the ‘ j ’ th household is

⁷The 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:
 $\ln C^j(u, P) = a_0^j + \sum_{i=1}^M a_i^j \ln P_i + \left(\frac{1}{2}\right) \sum_{i=1}^M \sum_{k=1}^M a_{ik}^j \ln P_i \ln P_k + b_0^j \ln u^j + \sum_{i=1}^M b_i^j \ln P_i \ln u^j + \left(\frac{1}{2}\right) b_{00} (\ln u^j)^2$. The parameters satisfy the following restrictions: $a_{ik}^j = a_{ki}^j \quad \forall i = 1, 2, \dots, M, k = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$; $\sum_{i=1}^M a_i^j = 1$; $\sum_{i=1}^M b_i^j = 0$; $\sum_{k=1}^M a_{ik}^j = 0 \quad \forall i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$.

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

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.⁸ 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

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

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

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

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.

⁸As we shall see, this makes the bias expression interpretable in terms of relative prices.

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

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

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

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

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

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

Hence, the democratic aggregate COLI is defined by

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

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

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

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

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

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

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

where the expectations operator is indexed by p to remind us that the averaging is plutocratic.

The corresponding plutocratic representative agent index is

$$(4) \quad R_p = T(E_p(\mathbf{s}))$$

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

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

This can be decomposed into two terms as follows⁹,

(5)

$$g = \frac{E_d(T(\mathbf{s})) - T(E_d(\mathbf{s}))}{T(E_p(\mathbf{s}))} + \frac{T(E_d(\mathbf{s})) - T(E_p(\mathbf{s}))}{T(E_p(\mathbf{s}))}$$

The second term on the right-hand side of (5) arises because of 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 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

⁹ The expression of the bias reported in equation (5) needs to be multiplied by 100 to get the percentage figures.

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

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

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

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 and 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

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

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

$$(6) \quad g_1 \equiv \frac{E_d[T(\mathbf{s})] - T[E_d(\mathbf{s})]}{T[E_d(\mathbf{s})]} \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]} \approx \left(\frac{1}{2}\right) \text{var}_d \left[\sum_{i=1}^{M-1} s_i \ln \lambda_i \right] \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]}$$

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 to this paper. The expression in (6) is clearly non-negative. 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 $\lambda_i = 1 \forall i = 1, 2, \dots, M - 1$.¹⁰ 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

$$(7) \quad g_2 \approx \left(\frac{1}{2}\right) var_d \left[\sum_{i=1}^{M-1} \alpha_i \ln \lambda_i \right] \frac{C(E_d(\alpha))}{C(E_p(\alpha))}$$

where α_i is the fixed budget share of ‘ i ’th 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.¹¹

The curvature bias component closely resembles the difference between a Carli and Jevons index derived by Diewert (2004). In the aggregation of price ratios, a 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.

¹⁰ Recall that the percentage change in the relative price of the ‘ i ’th commodity is given by $(\lambda_i - 1)$.

¹¹Cobb–Douglas price index is also called Geometric Laspeyres Index (Balk 2009). It can be considered as a geometric version of the Lowe index with updated weights from a third period.

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.

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 measures of dispersion – both evaluated by the democratic density. 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.¹²

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 2004–05 and 2011–12.¹³ 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.¹⁴ The prices of these categories are assumed to increase at a rate of 80%. Prices of all other

¹²In 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.* (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.

¹³For 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.

¹⁴These are the food categories of ‘cereals and cereal substitutes’ and ‘pulse and pulse products’.

categories are assumed to increase at a rate of 20% (including miscellaneous non-food). 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 Table 3 and Table 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.

United States

In the United States, the consumer expenditure survey is conducted by the Bureau of Labor Statistics. We use the Quarterly Interview Survey for the period 2015-18. 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.¹⁵

¹⁵Any household/consumer unit is asked to report their expenditure for the last three 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 don't 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

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 2018 and two hypothetical scenarios.¹⁶ 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 Bureau of Labor Statistics (BLS). The first column of Table 6 reports the observed price change when aggregated into 8 categories. The scenario in the second column assumes that all commodities within the food and beverages, shelter and utilities 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 3 rows report the bias calculations.

As expected, the plutocratic bias reverses in sign between the second and the third scenario. 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.

each consumer unit, the expenditure made in the last and current quarter needs to be added. It is also called the 'collection period' expenditure.

¹⁶ 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.

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 household surveys in India which are cross-sectional. Panels can be constructed from the US quarterly interview survey where households are surveyed for four quarters before they are rotated out. We construct four panels for the adjacent years 2015-16. 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 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 is not commonly available. In the United States, for instance, the quarterly interview survey 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 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 since 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 Tornqvist index can be expressed as

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

where the budget shares are averages over 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

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

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

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

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

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

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 2004/05 and 2011/12, i.e., 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 second quarter of 2018 (called scenario 1 in Table 6).¹⁷For this period, panel data was 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 comes 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 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

¹⁷ The estimates are similar for other quarters.

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 US data. An upper bound to the bias for Tornqvist index can be found from cross-sectional data alone which is otherwise insufficient to estimate the exact aggregation bias corresponding to 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 2004-05 and 2011-12, 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 US 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 Tornqvist representative agent bias computed from the US 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. Secondly, 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.

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. The mean budget shares as well as measures of dispersion are evaluated by democratic density. CV stands for 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. The mean budget shares as well as measures of dispersion are evaluated by democratic density. CV stands for Coefficient of Variation.

Table 3: Change in Prices and Representative Agent Bias for Cobb-Douglas Index (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
Curvature bias (in %)	0.06	0.07	0.07
Plutocratic Bias (in %)	0.47	1.22	-1.21
Overall Bias (in %)=Curvature bias (in %) + Plutocratic Bias (in %)	0.53	1.29	-1.14

Note: In scenario 1, we consider the observed change in prices for all categories between 2011–12 and 2004–05. 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% . .

Table 4: Change in Prices and Representative Agent Bias for Cobb-Douglas Index (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
Curvature bias	0.05	0.06	0.06
Plutocratic Bias (in %)	1.33	1.18	-1.17
Overall Bias (in %)=Curvature bias (in %) + Plutocratic Bias (in %)	1.38	1.24	-1.11

Note: In scenario 1, we consider the observed change in prices for all categories between 2011–12 and 2004–05. 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%.

Table 5: Budget Share of Commodities (US Data: 2nd Quarter, 2015)

Commodity	Mean	Std. Dev.	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 second quarter of the Quarterly Interview Survey, 2015. The figures for the other quarters are almost same. Each of these categories has many sub-categories. The budget shares of these sub-categories 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 8 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 Coefficient of Variation.

Table 6: Change in Prices and Representative Agent Bias for Cobb-Douglas Index (US Data: Quarterly Interview Survey)

Changes in Prices (in %)	Scenario1	Scenario2	Scenario3
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(in %)	0.02	0.005	0.005
Plutocratic Bias (in %)	0.51	0.38	-0.38
Overall Bias (in %) =Curvature bias (in %) + Plutocratic Bias (in %)	0.53	0.385	-0.375

Note:-The price changes have been computed from the price indices of individual commodities/categories available from Bureau of Labor Statistics. There are 41 categories used in computation of the representative agent bias. But the above table aggregates them into 8 categories by taking the geometric average of the change in prices of the individual sub-categories under each of these 8 categories. Scenario 1 corresponds to the actual change in prices between second quarter of 2018 and 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.

Table 7: Change in Prices and Representative Agent Bias for Tornqvist Index (US Data: Quarterly Interview Survey)

Changes in Prices (in %)	Panel1	Panel2	Panel3	Panel4
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 (in %)	0.002	0.003	0.006	0.003
Plutocratic Bias (in %)	-0.072	0.009	-0.058	-0.027
Overall Bias (in %) =Curvature bias (in %) + Plutocratic Bias (in %)	-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 Bureau of Labor Statistics. There are 41 categories used in computation of the representative agent bias. But the above table aggregates them into 8 categories by taking the geometric average of the change in prices of the individual sub-categories under each of these 8 categories. The price changes for any panel are the change in prices between the quarters that constitute the panel. Any negative price change implies decline in the price of that commodity/category. Panel 1 is constituted of those households who are interviewed both in first and fourth quarter of 2015. Similarly, in panel 2 we have households who are interviewed both in the second quarter of 2015 and 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.

Table 8: Upper Bound (in %) to the Representative Agent Bias for the Tornqvist Index: India

	Upper Bound on Curvature Bias (in %)	Plutocratic Bias (in %)	Upper Bound on Overall Bias (in %)
Rural	0.08	0.53	0.61
Urban	0.14	1.3	1.44

Note: The calculations are based on the National Sample Survey (2004-05 and 2011-12). These upper bounds correspond to the observed price changes between 2011-12 and 2004-05 i.e. price changes under scenario 1 in Table 3 and Table 4.

Table 9: Upper Bound (in %) to the Representative Agent Bias for the Tornqvist Index: US

	Upper Bound on Curvature Bias (in %)	Plutocratic Bias (in %)	Upper Bound on Overall Bias (in %)
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 US 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 second quarter of 2015 i.e. 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 second quarter of 2018 and second quarter of 2015, we can't 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.

Appendix

1. Proof of Proposition1

Convexity of $T(\mathbf{s})$ requires the matrix of the second derivative of $T(\mathbf{s})$, i.e. the Hessian matrix, to be positive semidefinite. A diagonal element of the matrix is $\frac{\partial^2 T}{\partial s_i^2} = T \cdot [\ln \lambda_i^2] \forall i =$

$1, 2, \dots, M - 1$, where T is the Tornqvist index. An off-diagonal element can be written as

$$\frac{\partial^2 T}{\partial s_i \partial s_k} = T \cdot [\ln \lambda_i \ln \lambda_k] \quad \forall i = 1, 2, \dots, M - 1, k = 1, 2, \dots, M - 1; i \neq k.$$

Hence the Hessian matrix can be written as

$$H = T(D \cdot D^t)$$

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

$$Y^t H Y = Y^t T(D \cdot D^t) Y = T(Y^t D \cdot D^t Y) = T((D^t Y)^t (D^t Y)) = T \|D^t Y\|^2 \geq 0$$

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

2. Proof of Proposition3

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

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

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 \mathbf{h} be the vector $(h_1 h_2 \dots h_{M-1})$. Let

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

It can be shown that $R_2(\mathbf{E}_d(\mathbf{s}), \mathbf{h})$ is $o(\|\mathbf{h}\|^2)$, i.e. $\frac{R_2(\mathbf{E}_d(\mathbf{s}), \mathbf{h})}{\|\mathbf{h}\|^2}$ tends to zero as \mathbf{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

$$\begin{aligned}
E_d[T(\mathbf{s})] - T[E_d(\mathbf{s})] &\approx \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} E_d[s_i - E_d(s_i)]^2 \left(\frac{\partial^2 T}{\partial s_i^2}\right) \\
&+ \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} E_d[(s_i - E_d(s_i))(s_k - E_d(s_k))] \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right) \\
\text{(A2)} \quad &= \left(\frac{1}{2}\right) [\sum_{i=1}^{M-1} \text{var}_d(s_i) \left(\frac{\partial^2 T}{\partial s_i^2}\right) + \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} \text{cov}_d(s_i, s_k) \left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)]
\end{aligned}$$

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

$$\begin{aligned}
\frac{E_d[T(\mathbf{s})] - T[E_d(\mathbf{s})]}{T[E_d(\mathbf{s})]} &\approx \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \text{var}_d(s_i) \frac{\left(\frac{\partial^2 T}{\partial s_i^2}\right)}{T[E_d(\mathbf{s})]} \\
\text{(A3)} \quad &+ \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} \text{cov}_d(s_i, s_k) \frac{\left(\frac{\partial^2 T}{\partial s_i \partial s_k}\right)}{T[E_d(\mathbf{s})]}
\end{aligned}$$

Now, $\frac{\left(\frac{\partial^2 T}{\partial s_i^2}\right)}{T[E_d(\mathbf{s})]} = (\ln \lambda_i)^2 \forall i = 1, 2, \dots, M-1$ and $\frac{\frac{\partial^2 T}{\partial s_i \partial s_k}}{T[E_d(\mathbf{s})]} = (\ln \lambda_i)(\ln \lambda_k) \forall i = 1, 2, \dots, M-1$

$1; k = 1, 2, \dots, M-1; i \neq k$

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

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

$$= \left(\frac{1}{2}\right) \text{var}_d \left[\sum_{i=1}^{M-1} s_i \ln \lambda_i \right]$$

Therefore, the curvature bias is characterized by

$$(A4) \quad g_1 \approx \left(\frac{1}{2}\right) \text{var}_d \left[\sum_{i=1}^{M-1} s_i \ln \lambda_i \right] \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]}$$

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

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

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

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

It can be readily checked that $\frac{\|\mathbf{h}\|^3}{3!} U$ is $o(\|\mathbf{h}\|^2)$, i.e. dividing $\frac{\|\mathbf{h}\|^3}{3!} U$ by $\|\mathbf{h}\|^2$, we get

$\frac{\|\mathbf{h}\|}{3!} U$ and this goes to zero as $\mathbf{h} \rightarrow 0$ (provided that U is a positive constant). As $\frac{\|\mathbf{h}\|^3}{3!} U$ is

$o(\|\mathbf{h}\|^2)$ and $R_2(E(\mathbf{s}), \mathbf{h}) \leq \frac{\|\mathbf{h}\|^3}{3!} U$, $R_2(E(\mathbf{s}), \mathbf{h})$ is $o(\|\mathbf{h}\|^2)$ as well, i.e. $\frac{R_2(E(\mathbf{s}), \mathbf{h})}{\|\mathbf{h}\|^2}$ tends to

zero as \mathbf{h} tends to zero.

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

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

Now,

$$\begin{aligned} & |(\frac{\partial}{\partial s_1} + \frac{\partial}{\partial s_2} + \dots + \frac{\partial}{\partial s_{M-1}})^3 T(\mathbf{t})| \\ &= | \sum_{i=1}^{M-1} \frac{\partial^3 T(\mathbf{t})}{\partial s_i^3} + 3 \sum_{i=1}^{M-1} \sum_{k=1; i \neq k}^{M-1} \frac{\partial^3 T(\mathbf{t})}{\partial s_i^2 \partial s_k} \\ & \quad + \sum_{i=1}^{M-1} \sum_{k=1}^{M-1} \sum_{l=1; i \neq k \neq l}^{M-1} \frac{\partial^3 T(\mathbf{t})}{\partial s_i \partial s_k \partial s_l} | \end{aligned}$$

Since we are considering the absolute value of the derivative, $(\frac{\partial}{\partial s_1} + \frac{\partial}{\partial s_2} + \dots + \frac{\partial}{\partial s_{M-1}})^3 T(\mathbf{t})$, 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 curvature bias for the Tornqvist index is expressed as

$$\begin{aligned} (A5) \quad g_1 &\approx [(\frac{1}{2}) \sum_{i=1}^{M-1} var_d(s_i) (\ln \lambda_i)^2 \\ & \quad + (\frac{1}{2}) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} cov_d(s_i s_k) (\ln \lambda_i) (\ln \lambda_k)] \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]} \\ s_i &= (\frac{1}{2}) (s_i^1 + s_i^0); s_k = (\frac{1}{2}) (s_k^1 + s_k^0) \end{aligned}$$

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 $\frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]}$ 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

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

Now,

$$\begin{aligned} \text{var}_d(s_i) &= \text{var}_d\left(\frac{s_i^1 + s_i^0}{2}\right) = \left(\frac{1}{4}\right) \text{var}_d(s_i^1 + s_i^0) \\ &= \left(\frac{1}{4}\right) [\text{var}_d(s_i^1) + \text{var}_d(s_i^0) + 2\text{cov}_d(s_i^1, s_i^0)] \end{aligned}$$

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

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

Now,

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

where R_i^2 is the squared correlation coefficient between s_i^1 and $s_i^0 \forall i = 1, 2, \dots, M -$

1. Replacing $\frac{\text{cov}_d(s_i^1, s_i^0)}{\sqrt{\text{var}_d(s_i^1)\text{var}_d(s_i^0)}}$ by $(R_i^2)^{\left(\frac{1}{2}\right)}$, we can write down the variance as

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

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:

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

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

$$\begin{aligned} B1 &= \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \text{var}_d(s_i) (\ln \lambda_i)^2 \\ (A6) \quad &\leq \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \left[\text{var}_d(s_i^1) + \text{var}_d(s_i^0) + 2\sqrt{\text{var}_d(s_i^1)\text{var}_d(s_i^0)} \right] (\ln \lambda_i)^2 \end{aligned}$$

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

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

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

$$\begin{aligned} \text{cov}_d(s_i, s_k) &= \text{cov}_d((s_i^1 + s_i^0)/2, (s_k^1 + s_k^0)/2) \\ &= (1/4)[\text{cov}_d(s_i^0, s_k^0) + \text{cov}_d(s_i^1, s_k^0) + \text{cov}_d(s_i^0, s_k^1) + \text{cov}_d(s_i^1, s_k^1)] \end{aligned}$$

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

$$\text{cov}_d(s_i^1, s_k^0) \leq \sqrt{\text{var}_d(s_i^1)\text{var}_d(s_k^0)} \text{ and } \text{cov}_d(s_i^0, s_k^1) \leq \sqrt{\text{var}_d(s_i^0)\text{var}_d(s_k^1)}$$

Therefore, upper bound on the entire covariance term $B2$ can be written as

$$\begin{aligned} \text{(A7)} \quad & \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} \text{cov}_d(s_i s_k) (\ln \lambda_i)(\ln \lambda_k) \\ & \leq \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} [\text{cov}_d(s_i^0, s_k^0) + \text{cov}_d(s_i^1, s_k^1)] (\ln \lambda_i)(\ln \lambda_k) \\ & \quad + \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} \left[\sqrt{\text{var}_d(s_i^1)\text{var}_d(s_k^0)} \right. \\ & \quad \left. + \sqrt{\text{var}_d(s_i^0)\text{var}_d(s_k^1)} \right] (\ln \lambda_i)(\ln \lambda_k) \end{aligned}$$

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

$$\begin{aligned} \text{(A8)} \quad g_1 & \approx \left[\left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \text{var}_d(s_i) (\ln \lambda_i)^2 + \left(\frac{1}{2}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} \text{cov}_d(s_i, s_k) (\ln \lambda_i)(\ln \lambda_k) \right] \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]} \\ & \leq \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \left[\text{var}_d(s_i^1) + \text{var}_d(s_i^0) + 2\sqrt{\text{var}_d(s_i^1)\text{var}_d(s_i^0)} \right] (\ln \lambda_i)^2 \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]} \\ & \quad + \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} [\text{cov}_d(s_i^0, s_k^0) + \text{cov}_d(s_i^1, s_k^1)] (\ln \lambda_i)(\ln \lambda_k) \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]} \\ & \quad + \left(\frac{1}{8}\right) \sum_{i=1}^{M-1} \sum_{k=1, i \neq k}^{M-1} \left[\sqrt{\text{var}_d(s_i^1)\text{var}_d(s_k^0)} \right. \\ & \quad \left. + \sqrt{\text{var}_d(s_i^0)\text{var}_d(s_k^1)} \right] (\ln \lambda_i)(\ln \lambda_k) \frac{T[E_d(\mathbf{s})]}{T[E_p(\mathbf{s})]} \end{aligned}$$

The upper bound on the curvature bias i.e. the right hand side of the 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.

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.
- Balk, B. M. (2009). 'Lowe and Cobb-Douglas CPIs and their substitution bias'. Proceedings of the Eleventh Meeting of the International Working Group on Price Indices (Neuchâtel, Switzerland). Available on the Internet from www.ottawagroup.org.
- 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., and 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 Economía Aplicada*, **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.
- Diewert, W. Erwin (2004), 'Basic Index Number Theory', in Paul A. Arm-knecht (ed.), *Producer Price Index Manual: Theory and Practice (PPI Manual)*, published by the International Labour Office, International Monetary Fund, Organisation for Economic Co-operation and Development, Eurostat, United Nations, and The World Bank.

- 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 Mary F. Kokoski. (1996), 'An Experimental Consumer Price Index for the Poor,' *Monthly Labor Review*, 119:9, pp. 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.
- Kaplan, G. and Schulhofer-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 and Measuring 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