# DISCUSSION PAPER SERIES

DP16866


## Nonparametric Measurement of Long-Run Growth in Consumer Welfare


Xavier Jaravel and Danial Lashkari



**MACROECONOMICS AND GROWTH**

**PUBLIC ECONOMICS**

ISSN 0265-8003

# Nonparametric Measurement of Long-Run Growth in Consumer Welfare

*Xavier Jaravel and Danial Lashkari*

Discussion Paper DP16866
Published 08 January 2022
Submitted 29 December 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

# Nonparametric Measurement of Long-Run Growth in Consumer Welfare


## Abstract

How should we measure long-run changes in consumer welfare? This paper proposes a nonparametric approach with arbitrary observable preference heterogeneity, e.g. when expenditure shares vary with income. Our approach only requires data on the consumption baskets of a cross-section of consumers facing a common set of prices. We take nominal expenditures under a constant set of prices as our money-metric for real consumption (welfare), and we derive a consistent measure of its growth in terms of a correction to the conventional quantity indices. Our correction accounts for the cross-sectional dependence of price indices on consumer income and any obserable source of preference heterogeneity. We use nonparametric methods to approximate these corrections and provide bounds on the resulting approximation errors that hold under arbitrary underlying preferences. Applying our methodology to the measurement of growth in real consumption per capita in the United States, we find a sizable correction to standard measures of growth in the postwar era–a period of fast growth combined with substantial inflation gaps across income groups



JEL Classification: N/A

Keywords: N/A



Xavier Jaravel - xavier.jaravel@gmail.com
*LSE and CEPR*

Danial Lashkari - danial.lashkari@bc.edu
*Boston College*

# Nonparametric Measurement of Long-Run Growth in Consumer Welfare

Xavier Jaravel[†] Danial Lashkari[‡]

December 21, 2021

**Abstract**

How should we measure long-run changes in consumer welfare? This paper proposes a nonparametric approach with arbitrary observable preference heterogeneity, e.g. when expenditure shares vary with income. Our approach only requires data on the consumption baskets of a cross-section of consumers facing a common set of prices. We take nominal expenditures under a constant set of prices as our money-metric for real consumption (welfare), and we derive a consistent measure of its growth in terms of a correction to the conventional quantity indices. Our correction accounts for the cross-sectional dependence of price indices on consumer income and any obserable source of preference heterogeneity. We use nonparametric methods to approximate these corrections and provide bounds on the resulting approximation errors that hold under arbitrary underlying preferences. Applying our methodology to the measurement of growth in real consumption per capita in the United States, we find a sizable correction to standard measures of growth in the postwar era–a period of fast growth combined with substantial inflation gaps across income groups.

We thank David Baqaee, Ariel Burstein, Ryan Chahrour, Shakeeb Khan, Jaromir Nosal, and Robert Ulbricht for stimulating conversations and helpful feedback We are grateful to Sylvia Tian for excellent research assistance.

[†]London School of Economics. Email: x.jaravel@lse.ac.uk . Website: https://www.xavierjaravel.com.

[‡]Boston College. Email: danial.lashkari@bc.edu. Website: https://www.daniallashkari.com.

# 1 Introduction

How should we measure long-run changes in living standards? Classical demand theory shows that intuitive index numbers, which simply aggregate observed changes in consumed quantities and prices, may provide precise measures of the change in consumer welfare. However, this powerful insight requires the crucial assumption that the composition of demand remains independent of consumer income. This so-called homotheticity assumption runs counter to the empirical regularity that demand for many goods such as food or housing systematically depends on income, a fact known since at least Engel (1857). It also belies the growing empirical evidence on sizable and systematic differences in inflation rates across income groups, typically featuring lower inflation rates for higher-income groups in the United States (e.g., Kaplan and Schulhofer-Wohl, 2017; Jaravel, 2019; Argente and Lee, 2021).

Despite this important and well-known theoretical limitation, classical price indices remain widely used in practice due to their simplicity, flexibility, and generality. Little is known about potential biases arising from the restrictive homotheticity assumption in the resulting measures of long-run growth in average standards of living. Current alternatives require us to take a stance on and estimate the structure of the demand system–a task that is not feasible with many available datasets. For instance, Baqaee and Burstein (2021) have recently offered an approach that relies on the knowledge of the elasticities of substitution across goods to construct measures of welfare growth (see also Samuelson and Swamy, 1974).

In this paper, we develop a new approach for measuring welfare change that allows for flexible dependence of the patterns of demand on income and other sources of observed heterogeneity. Compared to the standard index number theory, the only additional requirement is access to a cross-section of consumed quantities under common prices in at least one period; such data is widely available through standard surveys of consumption expenditure. Our approach nonparametrically estimates the cross-sectional dependence of price indices on consumer income. We show that the knowledge of the degree of this dependence is sufficient to provide precise approximations for a theoretically consistent measure of real consumption, while imposing minimal restrictions on the underlying preferences. This approach remains valid under any observable heterogeneity. We then apply our method to account for nonhomotheticity of demand in measuring long-run growth in the welfare of average US consumer from 1954 to 2019. In addition to improving the measurement of long-run growth and inflation inequality, our new approach can have important policy implications, such as the indexation of the poverty line and a more efficient targeting of welfare benefits.

We begin by defining real consumption indices, expressing the expenditure required to achieve initial and final levels of welfare under a constant set of base prices, and cost-of-living indices, expressing the expenditure required to achieve a base level of welfare under initial and final sets of prices. We contrast these *structural* indices, defined in terms of the unobserved underlying preferences, from *reduced-form* indices such as the geometric index and Törnqvist index, defined solely in terms of observed initial and final prices and expenditures. Using this distinction, we first restate the classical results under the assumption of preference homotheticity, by establishing bounds on the approximation error when we use reduced-form indices to approximate the structural indices. We then partially generalize these results to the nonhomothetic case when the structural indices are expressed from the perspective of a base period between initial and final periods.[1] This analysis leads to an important insight: under nonhomotheticity, the choice of the base period for expressing real consumption and cost-of-living indices may lead to different measures of welfare growth.

To accurately measure welfare change over long time horizons, we need to maintain a constant base period as the basis for consistent comparisons while we consecutively approximate real consumption indices over multiple periods. We characterize how welfare comparisons in terms of a distant base period deviate from those under the current base period using a *nonhomotheticity correction function*. The correction accounts for the covariance across goods between income elasticities and the cumulative price inflation between the base and current periods. As we move away from the base period and the cumulative inflation becomes large, the correction may grow to become first-order in welfare change.

To see the intuition behind this correction, consider a setting where consumer welfare is rising over a time horizon in which price inflation rates are relatively lower for goods with high income elasticity (luxuries). Let us express real consumption in terms of the prices in the initial period, such that in this period total consumer expenditure is linear in (and identical to) real consumption. As time passes, the *relative* cost of achieving higher levels of real consumption falls, since the goods consumed by richer consumers are getting relatively less expensive. In other words, consumer expenditure as a function real consumption becomes more and more concave over time. Thus, a given rise in total expenditure translates into increasingly larger gains in real consumption as consumers become richer. Assuming homotheticity, the conventional approach ignores this gradual fall in the curvature of the expenditure function and underestimates the growth of real consumption in terms of the initial base period. If we instead express real consumption in terms of the expenditure in the final period, the same logic implies that conventional indices overestimate the growth in all preceding periods.[2] Our nonhomotheticity correction measures the

[1] This result is similar in spirit to the results of Diewert (1976), who shows that the Törnqvist price indices coincide with the change in a nonhomothetic translog expenditure functions for an intermediate level of utility between the initial and the final periods.

[2] In this case, since total consumer expenditure is identical to real consumption in the final period, it must be a convex function of real consumption in all prior periods. This leads to overestimating the growth of real consump-

contribution of changes in the curvature of the expenditure function to account for its effect on measured growth in terms of any base period of interest.

Our main contribution is to provide simple procedures to approximate the nonhomotheticity correction function based on cross-sectional data under arbitrary nonhomothetic preferences.[3] The core idea is simple and intuitive. In any given base period, total nominal consumer expenditure by definition coincides with real consumption. This allows us to nonparametrically approximate the elasticity of the expenditure function with respect to real consumption using the cross-sectional variations across consumers in reduced-form price indices. Using this elasticity, we can infer the nonomotheticity correction and approximate real consumption in periods immediately away from the base period. We can then recursively apply the same strategy in the subsequent periods as we progress in time away from the base period to approximate real consumption over our entire period of interest. We provide two such approximations, first and second order in annual inflation and total nominal expenditure growth, and provide bounds on the error in each case using standard results in nonparametric function approximation literature.[4]

We study the accuracy of our strategy in measuring long-run growth and inequality using an illustrative example with known preference parameters. We consider the nonhomothetic CES (nhCES) preferences estimated for three main categories of sectoral goods, agriculture, manufacturing, and services, by Comin et al. (2021). We generate data on consumption choices in a synthetic sample of households, calibrating the average inflation and growth in nominal expenditure to those in the US in the 1953-2019 period. We simulate the data under different values for the covariance between price inflation rates and income elasticities across goods. Knowing the underlying expenditure function, we can compute the structural indices and compare them against the conventional reduced-form indices and those that reflect our nonhomotheticity correction. As predicted by our theory, high levels of covariance, when combined with high growth in real consumption, lead to sizable errors in the conventional reduced-form price indices over long time horizons. We confirm that our procedure accurately recovers the evolution of the exact index using the observed cross-sectional data and without *any* knowledge of the underlying preference parameters.

We further apply our approach to real data from the US and quantify the magnitude of the bias due to the observed variations in reduced-form price indices in the cross-section of households. We build a dataset linking income-specific expenditure shares (five quintiles of income) from the

tion when using the final period as base.

[3]This result extends more broadly to setting with arbitrary observed heterogeneity, but it rules out unobserved heterogeneity.

[4]Despite its larger approximation error, the first-order approximation has the advantage that it requires access to a cross-section of consumers only in the base year. In contrast, the second-order approximation requires access to a cross-section of consumers overlapping every consecutive periods of interest.

Consumer Expenditure Survey (CEX) to inflation rates from the Consumer Price Index (CPI) across product categories covering the full consumption baskets of US households from 1984 to 2019. We complement this data with information on CPI inflation rates and on aggregate expenditure shares from the Bureau of Labor Statistics (BLS) over the longer horizon of 1954-2019 period.

We first focus on the period 1984-2019 where cross-sectional data is available. In line with the work cited above, using conventional uncorrected indices we find lower inflation and higher real consumption growth for high-income US households. As with the example discussed above, reduced-form indices underestimate growth in real consumption when expressed with the initial base (1984), and overestimate it when expressed with the final base (2019). We find that the magnitude of the bias in the uncorrected measure of annual growth of real consumption rises as we move away from the base period, reaching around 8% of the measured annual growth for the highest quintile. Nevertheless, since over this period growth is modest, the overall bias in uncorrected cumulative growth, both for the mean and for the highest quintile, reaches only around 1 percentage point. We find slightly smaller magnitudes for the bias when we rely on less disaggregated data from the BLS, and slightly higher magnitudes when we consider a smaller sample of more disaggregated grocery products using Nielsen's scanner-level data (2004-2014).

Examining the earlier postwar US experience may be of particular interest due to is substantially higher rates of real consumption growth. Using the earliest available cross-sectional data in 1984 as the basis of our correction, we extend the analysis to the entire 1954-2019 period. In terms of the cumulative growth over the entire period, the upward bias is around 14 percentage points, from a benchmark of 141 percentage points in measured cumulative growth. The bulk of this bias is coming from the 1954-1984 period featuring faster growth: the cumulative overestimation over this 30-year period is around 10 percentage points (or 32 basis points annually). When real consumption is expressed in terms of 1984 prices, the bias in the uncorrected measurement of growth is smaller over the entire period (1954 to 2019). This is because, just like the example discussed above, the direction of the bias changes sign over the periods before (1954-1984) and after (1984-2019) the base year. Nevertheless, even using 1984 prices as base, welfare comparisons using the conventional reduced-form indices remain subject to error. For instance, the cumulative growth in real consumption from 1954 to 1984 is overestimated by around 4 percentage point. We conclude that accounting for nonhomotheticity can matter for consistent quantitative comparisons of welfare over long time horizons.

In the last part of the paper, we generalize our results to setting that involve heterogeneity, in addition to income, in other observed consumer characteristics that potentially evolve over time. We define generalized notions of structural quantity and price indices and provide nonparametric estimation strategies that allow us to approximate them based on cross-sectional variations in reduced-form prices indices as functions of observables.

**Prior Work** Our paper builds on and contributes to three strands of literature. First, we extend the literature on index number theory (e.g., Pollak, 1990; Diewert, 1993), which has enabled transparent and consistent comparisons of the aggregate measures of consumption and production over time and space only relying on observables under the assumption of homotheticity. As emphasized by Samuelson and Swamy (e.g., 1974), many classical results do not generalize beyond settings involving homotheticity in preferences. Under nonhomotheticity, Diewert (1976) has showed that one can still rely on the conventional price indices to measure changes in welfare locally. However, as we show here, these results do not generalize to welfare comparisons over long time horizons. We provide a detailed comparisons of our results from those in the classical theory in Appendix A.2.

Second, we advance a growing literature raising the point that standard price indices suffer from a bias due to nonhomotheticities, whose magnitude relates to the covariance between income elasticities and price changes (e.g., Fajgelbaum and Khandelwal, 2016; Atkin et al., 2020; Baqaee and Burstein, 2021). In particular, Baqaee and Burstein (2021) have recently highlighted the failure of standard divisia indices to measure welfare-relevant measures of growth in real consumption. They suggest relying on the estimates of the elasticities of substitution to account for the role of nonhomotheticity.[5] In contrast, we provide a nonparametric approach that does not require specifying the underlying demand functions. The importance of the covariance between income elasticities and inflation for measuring welfare change is also noted by Fajgelbaum and Khandelwal (2016) and Atkin et al. (2020). Fajgelbaum and Khandelwal (2016) measure changes in welfare gains from trade liberalization across different income groups in a parametric setting and under the assumption of an AIDS demand system (Deaton and Muellbauer, 1980). Atkin et al. (2020) consider the problem of welfare measurement in the absence of reliable price data, and rely on separability assumptions on the structure of preferences to infer welfare from shifts in the Engel curves. For this procedure to hold without the need for estimation of structural elasticities of substitution, they rule out exactly the covariance patterns that lead to large nonhomotheticity corrections in our framework.[6] In summary, while this literature provides parametric corrections for the bias, our contribution is to provide a nonparametric correction, which is valid under arbitrary preferences where all heterogeneity is in terms of observables.

Third, we contribute to the literature on the measurement of inflation inequality (e.g., Hobijn

[5] Baqaee and Burstein (2021) additionally study the consequences of the endogeneity of prices in general equilibrium, as well as unobserved heterogeneity, e.g. taste shocks. The latter effects have also been recently considered by Redding and Weinstein (2020).

[6] Atkin et al. (2020) also analyze the case when relative prices change, in which case their procedure requires computing (unobserved) compensated shifts in expenditure shares due to the change in relative prices (see their equation (2)).

and Lagakos, 2005; McGranahan and Paulson, 2006; Kaplan and Schulhofer-Wohl, 2017; Jaravel, 2019; Argente and Lee, 2021). Prior work on inflation inequality has posited the existence of separate homothetic indices for different income groups. We apply our methodology to provide estimates of inflation inequality that are robust to potential biases arising from nonhomotheticities. Using a new linked dataset covering the period 1953-2019 in the United States, we apply our methodology to the measurement of short, medium, and long run growth in real consumption, on average and across the income distribution, and we quantify the magnitude of the bias in the conventional measures in each case.

# 2 Welfare and Index Numbers Under Nonhomotheticity

In this section, we establish three results. First, we show that measures of welfare changes systematically vary with the choice of reference prices under non-homotheticities. We do so in Section 2.1, using "structural" price and quantity indices expressed in terms of the unobserved underlying preferences. We also show how these indices relate to the concepts of equivalent and compensating variations. Second, we show that nonhomotheticity requires us to hold constant the reference period to implement principled welfare comparisons over the long run; in Section 2.2, we discuss the implication of this result when chaining "reduced-form" price indices, which are computed using observed prices and expenditures only. Finally, in Section 2.3 we derive the bias induced by nonhomotheticity for the reduced-form price indices and express it in terms of the covariance between real income elasticities and price inflation across goods. Appendix A.2 provides a detailed discussion of the connection between the results of this section and the results of the classical index number theory.

## 2.1 Definitions

Consider a setting where we observe the composition of consumption expenditures for a consumer whose total expenditure evolves over time. We assume that the households face the same sequence of prices $\boldsymbol{p}_{(0,T)} \equiv (\boldsymbol{p}_t)_{t=0}^T$ for $I$ products, where $\mathbf{0} \ll \boldsymbol{p}_t \in \mathbb{R}^I$, and their consumption pattern of is characterized by the sequence of vector of quantities $\boldsymbol{q}_{(0,T)} \equiv (\boldsymbol{q}_t)_{t=0}^T$, where $\mathbf{0} \ll \boldsymbol{q}_t \in \mathbb{R}^I$. Correspondingly, we let $y_t$ and $s_{i,t} \equiv p_{i,t}q_{i,t}/y_t$ denote the total expenditure and the share of good $i$ in total expenditure at time $t$, respectively.

We first define conventional price and quantity indices, which we call *reduced-form* price indices since they are defined without reference to any underlying structure on consumer preferences and can be computed in terms of observed expenditures and prices.

**Definition 1.** *Reduced-Form Price and Quantity Indices.* A pair of conjugate price $\mathbb{P}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)$ and quantity $\mathbb{Q}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)$ indices are positive-valued functions of the initial and final vectors of prices and quantities that satisfy

$$\mathbb{P}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)\times\mathbb{Q}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)=\frac{y_t}{y_{t_0}}, \tag{1}$$

as well as $\mathbb{P}\left(\boldsymbol{p}_{t_0},\alpha\boldsymbol{p}_{t_0};\boldsymbol{q}_{t_0},\boldsymbol{q}_{t_0}\right)\equiv\alpha$ and $\mathbb{Q}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_{t_0};\boldsymbol{q}_{t_0},\alpha\boldsymbol{q}_{t_0}\right)\equiv\alpha$ for all $\alpha>0$ and for vectors $0\ll\boldsymbol{p}_{t_0},\boldsymbol{q}_{t_0},\boldsymbol{p}_t,\boldsymbol{q}_t\in\mathbb{R}^I$.

The price and quantity indices allow us to aggregate the changes in a vector of prices and quantities either in terms of a single price or quantity index without the knowledge of the underlying form of consumer preferences. The condition in Equation 1 ensures that conjugate price and quantity indices together decompose the growth in total expenditure into a component aggregating prices and another aggregating quantities.[7] For instance, in this paper, we focus on two reduced-form price indices, the geometric index and the Törqvist index, defined respectively as[8]

$$\mathbb{P}_G\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)\equiv\prod_{i=1}^{I}\left(\frac{p_{i,t}}{p_{i,t_0}}\right)^{s_{i,t_0}}, \tag{2}$$

$$\mathbb{P}_T\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)\equiv\prod_{i=1}^{I}\left(\frac{p_{i,t}}{p_{i,t_0}}\right)^{\frac{1}{2}\left(s_{i,t_0}+s_{i,t}\right)}, \tag{3}$$

which suitably lend themselves best to the measurement of the growth in prices. The corresponding conjugate quantity indices are defined using definitions (2), (3), and condition (1).

We next define another class of price and quantity indices that allow us to evaluate changes in welfare consumer in a theoretically consistent framework. We call these indices *structural,* since they are explicitly defined with reference to the underlying preferences of consumers choosing quantities given observed prices. Thus, we first need to make the following assumption on the underlying preferences that rationalize the observed choices of consumers.

---

[7] This condition is known as Fisher's "weak factor reversal" condition (Diewert, 1993).

[8] The geometric price index is a first-order price index in that it only relies on the basket of quantities in one terminal period to aggregate price changes. Other examples of first-order price indices include the Laspeyres and Paasche price indices defined as $\mathbb{P}_L\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)\equiv\boldsymbol{p}_t\cdot\boldsymbol{q}_{t_0}/y_{t_0}$ and $\mathbb{P}_P\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)\equiv y_t/\boldsymbol{p}_{t_0}\cdot\boldsymbol{q}_t$, respectively. These two indices simply account for the relative changes in the price of the initial and final basket of goods consumed. We can also define the Fisher price index as the geometric average of the Laspeyres and Paasche indices, i.e., $\mathbb{P}_F\equiv(\mathbb{P}_P\times\mathbb{P}_L)^{1/2}$. Just like the Törqvist index, the Fisher index is also a second-order index, relying on the basket of quantities in both terminal periods to aggregate price changes (other examples include the Sato-Vartia price index). For a comparison of the accuracy of first and second order indices, see Diewert (1978). The results below can be generalized to other first and second order indices including Laspeyres and Fisher indices, albeit with different bounds on the approximation errors.

**Assumption 1.** *The preferences of the consumer characterized by a utility function* $U(\cdot):\mathbb{R}^I \to \mathbb{R}$ *and a corresponding, third-order continuously differentiable expenditure function* $E(\cdot;\cdot)$, *such that* $u_t = U(\boldsymbol{q}_t)$ *is the utility at time* $t$, *the total expenditure of the household satisfies* $y_t = E(u_t;\boldsymbol{p}_t)$ *and we have:*

$$q_{i,t} = \frac{\partial E(u_t;\boldsymbol{p}_t)}{\partial p_{i,t}}, \qquad t \in \{0,\cdots,T\},\, i \in \{1,\cdots,I\}.$$

Relying on Assumption 1, we introduce two structural indices: the real consumption index and the the cost-of-living index. We first define these indices and then show how they provide a decomposition of changes in nominal expenditures into a quantity component and a price component.

**Definition 2.** *Structural Price and Quantity Indices.* Consider preferences characterized by utility and expenditure functions $U(\cdot)$ and $E(\cdot;\cdot)$, respectively. The cost-of-living (CoL) price index for a base vector of quantities $\boldsymbol{q}_b$ is defined as

$$\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right) \equiv \frac{E(U(\boldsymbol{q}_b);\boldsymbol{p}_t)}{E\left(U(\boldsymbol{q}_b);\boldsymbol{p}_{t_0}\right)}, \qquad 0 \le b, t_0, t \le T. \tag{4}$$

The real consumption (ReC) quantity index for a base vector of prices $\boldsymbol{p}_b$ is defined as[9]

$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_b\right) \equiv \frac{E(U(\boldsymbol{q}_t);\boldsymbol{p}_b)}{E\left(U\left(\boldsymbol{q}_{t_0}\right);\boldsymbol{p}_b\right)}, \qquad 0 \le b, t_0, t \le T. \tag{5}$$

The CoL index in Equation (4) has a straightforward economic interpretation: it captures the relative change in the total expenditure required for a consumer that would like to maintain utility at the level reached with the vector of base quantities $\boldsymbol{q}_b$, as prices change from the initial period $t_0$ to final period $t$. This index provides us with a price index, in the sense that it aggregates the inflation (or deflation) in the costs of achieving a certain level of welfare.

The ReC index in Equation (5) captures the relative change in the total expenditure for a consumer whose consumption basket changes from the initial period $\boldsymbol{q}_{t_0}$ to the final vector $\boldsymbol{q}_t$ of quantities, under the constant vector of base prices $\boldsymbol{p}_b$. This index provides us with a quantity index, in the sense that it aggregates the changes in the basket of quantities consumed in terms

[9] We can also define the CoL quantity index $\mathscr{Q}_{CL}(\boldsymbol{p}_0,\boldsymbol{p}_1;y_0,y_1;\boldsymbol{q}_b)$ and ReC price index $\mathscr{P}_{CL}(\boldsymbol{q}_0,\boldsymbol{q}_1;y_0,y_1;\boldsymbol{p}_b)$ using the corresponding weak factor reversal condition (1). Prior literature has not settled on a unified terminology for referring to the CoL and ReC indices defined here. Samuelson and Swamy (1974) refer to the CoL price and ReC quantity index as the economic price and quantity index. Diewert (1976) refers to the ReC quantity index as the Theil index of real income, whereas Diewert (1993) refers to CoL price index as the Konüs price index, to the CoL quantity index as the Konüs-Pollak quantity index, and to the ReC quantity and price indices as Allen indices. Pollak (1990) refers to the ReC quantity index as the preference field quantity index.

of the welfare of a consumer who faces a certain set of base prices. This index can thus provide a measure of economic growth using a money-metric.

**Welfare Comparisons over Two Periods** When we focus on a welfare comparison between one intial period $t_0$ and one final period $t$, we have two choices for the base vectors in our structural indices. It is straightforward to see that the ReC quantity index with the initial base prices is the conjugate of the CoL price index with the final base quantities in the sense of Equation (1). Therefore, both indices lead to the same measure of welfare change. Similarly, the ReC quantity index with the final base prices is the conjugate of the CoL price index with the initial base quantities, leading to an alternative measure of welfare change.

When comparing welfare between two periods, the two distinct measures of welfare change implied by the two conjugate pairs of structural indices correspond to the standard concepts of *equivalent and compensating variations*.[10] Comparisons based on the ReC index with the initial base prices and the CoL with the final base quantity correspond to the concept of equivalent variation. In this comparison, we consider the relative change in the initial level of total expenditure that is welfare-equivalent to the change in consumed quantities from the perspective of the consumer facing initial prices. Furthermore, comparisons based on the ReC index with the final base prices and the CoL with the initial base quantities correspond to the concept of compensating variation. In this comparison, we consider the relative change in the total expenditure that compensates the consumer for the change in prices from the initial to the final period in welfare terms. As is well-known, these two approaches to assessing welfare change do not necessarily coincide in general–a point to which we will return below.

**Welfare Comparisons Over Multiple Periods** When we consider welfare comparisons across multiple periods, we face many alternative choices for the base vectors in structural indices. The main appeal of the structural definition of the ReC index (in Definition 2) is that it leads to a theory-consistent money metric for welfare, defined as a specific monotonic transformation (cardinalization) $G_b(\cdot)$ of the utility function according to

$$q = G_b(u) \equiv E(u; \boldsymbol{p}_b), \tag{6}$$

[10] In this comparison, equivalent variation $EV$ is defined as the value satisfying $EV \equiv E\left(U(\boldsymbol{q}_t); \boldsymbol{p}_{t_0}\right) - E\left(U\left(\boldsymbol{q}_{t_0}\right); \boldsymbol{p}_{t_0}\right)$ whereas the compensating variation $CV$ is defined as $CV \equiv E(U(\boldsymbol{q}_t); \boldsymbol{p}_t) - E\left(U\left(\boldsymbol{q}_{t_0}\right); \boldsymbol{p}_t\right)$.

which we will henceforth refer to simply as *real consumption*. We can correspondingly define an expenditure function in terms of real consumption as

$$\widetilde{E}_b(q;\boldsymbol{p}) \equiv E\left(G_b^{-1}(q);\boldsymbol{p}\right), \tag{7}$$

and note that, by definition, $q \equiv \widetilde{E}_b(q;\boldsymbol{p}_b)$. From Equation (5), we find that the real consumption $q_t^b \equiv G_b(u_t)$ of consumer at time $t$ expressed in period-$b$ prices is related to the ReC index according to the relation $q_t^b = y_b \times \mathscr{Q}_{RC}(\boldsymbol{q}_b, \boldsymbol{q}_t; \boldsymbol{p}_b)$. Thus, real consumption, and the corresponding ReC index, change monotonically with utility of the consumer when expressed in a constant base vector of prices. In other words, we use the base prices to define a money metric for welfare, i.e. welfare is expressed in terms of prices from period $b$.

In contrast, the CoL price index with a constant base vector of quantities $b$ characterizes the expenditure needed to reach a certain standard of living, e.g., the poverty line, in all other periods. We can define a conjugate quantity index for the CoL price index using the condition in Equation (1) as $\mathscr{Q}_{CL} \equiv (y_t/y_{t_0})/\mathscr{P}_{CL}$. However, note that this alternative quantity index does not allow theory-consistent welfare comparisons across multiple periods.[11]

Lastly, we note that, when comparing across multiple periods, the structural price and quantity indices by definition satisfy the *chain principle* (Diewert, 1978):

$$\mathscr{P}_{CL}(\boldsymbol{p}_0, \boldsymbol{p}_T; \boldsymbol{q}_b) = \prod_{t=0}^{T-1} \mathscr{P}_{CL}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_b\right), \tag{8}$$

$$\mathscr{Q}_{RC}(\boldsymbol{q}_0, \boldsymbol{q}_T; \boldsymbol{p}_b) = \prod_{t=0}^{T-1} \mathscr{Q}_{RC}\left(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}; \boldsymbol{p}_b\right). \tag{9}$$

Crucially, the chain principle above generically holds only when the base vectors of prices and quantities are held constant. Thus, when we rely on the chain principle to accumulate the structural price and quantity indices over time, we need to use a single base period $b$, which will sometimes be outside of the two consecutive periods $t$ and $t+1$ within each link. In contrast, the chaining property does not necessarily hold for many standard reduced-form price and quantity indices. However, for each such reduced-form index, we can always define a chained index as follows

$$\mathbb{P}^c\left(\boldsymbol{p}_{(0,T)}; \boldsymbol{q}_{(0,T)}\right) \equiv \prod_{t=0}^{T-1} \mathbb{P}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t, \boldsymbol{q}_{t+1}\right), \tag{10}$$

[11] To demonstrate why, we formally define conjugate structural indices in Appendix A.1.1. In particular, the conjugate CoL quantity index from period $t_0$ to period $t$ with base quantities $b$ defined in Equation (A3) is the product of two ReC quantity indices: 1) the ReC quantity index from period $t_0$ to period $b$ with base prices $t_0$ and 2) the ReC quantity index from period $b$ to period $t$ with base prices $t$. Since the product involves two distinct notions of money metrics, it does not provide a well-defined money metric.

$$\mathbb{Q}^c\left(\boldsymbol{p}_{(0,T)};\boldsymbol{q}_{(0,T)}\right)\equiv\prod_{t=0}^{T-1}\mathbb{Q}\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t,\boldsymbol{q}_{t+1}\right). \tag{11}$$

## 2.2 Reduced-Form Indices as Sufficient Statistics for Welfare Change

Next, we characterize the relationship between the reduced-form price and quantity indices defined in Definition 1 and the structural ones defined in Definitions 2, which in fact enable meaningful welfare comparisons. In addition, we will rely on the following assumption throughout this section.

**Assumption 2.** *Assume that for $t \geq 0$, the maximum growth in prices in bounded above by a constant $\Delta_p < 1$ and the maximum growth in the total expenditure is bounded above by constant $\Delta_y < 1$, such that*

$$\max_{1\leq i\leq I,0\leq t\leq T}\left|\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)\right|\leq\Delta_p,\qquad \max_{0\leq t\leq T}\left|\log\left(\frac{y_{t+1}}{y_t}\right)\right|\leq\Delta_y.$$

Assumption 2 simply imposes bounds on the growth in prices and the total expenditures of households over time. An alternative way to think of this assumption is that prices and total expenditures are continuous functions of time and our data simply captures the observations of these functions over discrete time periods. Using this assumption, our key result is

**Lemma 1.** *Under Assumptions 1 and 2, if the utility function $U$ is homothetic, then for all vectors of base periods $b$, we have:*

$$\log\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right)=\log\mathbb{P}_G\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)+O\left(|t-t_0|^2\Delta_p^2\right), \tag{12}$$

$$=\log\mathbb{P}_T\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)+O\left(|t-t_0|^3\Delta_p^3\right). \tag{13}$$

*If the utility function $U$ is potentially nonhomothetic, and if the base period $b$ satisfies $t_0\leq b\leq t$, then we have:*

$$\log\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right)=\log\mathbb{P}_G\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)+O\left(|t-t_0|^2\left(\Delta_p+\Delta_y\right)^2\right). \tag{14}$$

*Finally, there exists a basket of goods $\boldsymbol{q}_b^*$ satisfying $\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_b^*;\boldsymbol{p}_b\right)=\mathscr{Q}_{RC}\left(\boldsymbol{q}_b^*,\boldsymbol{q}_t;\boldsymbol{p}_b\right)$ such that*

$$\log\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b^*\right)=\log\mathbb{P}_T\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t\right)+O\left(|t-t_0|^3\left(\Delta_p+\Delta_y\right)^3\right). \tag{15}$$

*Proof.* See Appendix A.3. □

Lemma 1 establishes the tight connection between reduced-form and structural indices under the assumption of preference homotheticity. It shows that the geometric and Törqvist price indices approximate the CoL price index to the first and second order of approximation for all base vectors. In fact, *the values of CoL price indices and the ReC quantity indices under preference homotheticity do not depend on the base vectors of quantities and prices considered.*[12] It then also follows that any CoL price index is the conjugate of any ReC quantity index regardless of the bases considered, in the sense of Equation (1).

Under homotheticity, we can rely on the chaining principle in Equations (8) and apply Lemma 1 recursively across consecutive periods $(t_0, t) \equiv (t, t+1)$ while keeping a constant base vector of quantities. Using the conjugacy of ReC quantity and CoL price indices, we can compute the ReC index over multiple peiods, e.g., from period 0 to $T$, to the second order of approximation in $\Delta_p$ as:

$$\log \mathscr{Q}_{RC}(\boldsymbol{q}_0, \boldsymbol{q}_T; \boldsymbol{p}_b) = \log\left(\frac{y_T/y_0}{\mathbb{P}^c_T\left(\boldsymbol{p}_{(0,T)}; \boldsymbol{q}_{(0,T)}\right)}\right) + T \cdot O\left(\Delta_p^3\right), \tag{16}$$

where $\mathbb{P}^c_T$ is the chained Törqvist price index defined by Equation (10). The first term on the right hand side of the equation above is of the order $T \times \left(\Delta_y + \Delta_p\right)$. Thus, the expression above shows that the standard chain formula can provide a fairly accurate measure of the ReC index over long time horizons.

Under nonhomothetic preferences, as we will discuss at length below, the values of structural indices depend on the choice of the base vectors of prices and quantities. Lemma 1 provides a generalization of the relation between reduced-form and structural indices to these case. If the base vector of quantities corresponds to a period *between the initial and the final periods*, the reduced-form price indices provide approximations of the CoL price index with errors that are comparable with those found under the homothetic case covered in Lemma 1.[13] The difficulty arises if we aim to apply the chain principle to evaluate welfare over longer horizons. In particular, we cannot apply Lemma (1) recursively across consecutive periods $(t_0, t) \equiv (t, t+1)$ from time 0 to $T$, because the base period $b$ in the lemma has to satisfy $t \leq b \leq t+1$. If we instead use the lemma over long horizons, e.g., letting $(t_0, t) \equiv (0, T)$ to allow for a constant base period, the error in the approximation features substitution bias and grows with the size of cumulative inflation/growth over the entire period, which could be substantial.[14]

[12] See Appendix A.1.2 for a detailed discussion.

[13] We can rely on the decomposition in Equation (A2) below to write a ReC quantity index with any base vector of quantities between the initial and the final periods in terms of corresponding CoL price indices, and rely again on Lemma 1 to approximate the ReC quantity index.

[14] These observations show that the seminal results of Diewert (1976), establishing that standard reduced-form price indices can be used for local welfare comparisons under non-homotheticities, do not generalize to welfare

In the next section, we show how to make nonparametric corrections to the standard reduced-form price indices to improve the approximation of the structural price and quantity indices.

## 2.3 The Nonhomotheticity Correction Function

When do reduced-form price indices fail to provide accurate approximations of the CoL price indices between periods $t_0$ and $t$ under nonhomothetic preferences? Lemma 1 tells us that when the basket of quantities is similar to that of the consumer in either of the two periods, the reduced-form price indices offer reasonable approximations. Thus, for this approximation to fail, the CoL price indices have to strongly depend on their base basket of quantities.

To examine the nature of this dependence, let us first use Shephard's lemma along with the definition of the expenditure function (7) to define the Hicksian demand function:

$$\Omega_i^b\left(q^b;\boldsymbol{p}_t\right) \equiv \frac{\partial \log \widetilde{E}_b\left(q^b;\boldsymbol{p}_t\right)}{\partial \log p_{it}}, \tag{17}$$

which characterizes the expenditure share for product $i$ under prices $\boldsymbol{p}_t$ for a consumer with real consumption $q$.[15] Let us also introduce a simplified notation to refer to the CoL price index that expresses how the cost-of-living changes from period $t_0$ to $t$ for a consumer who aims to maintain real consumption $U(\boldsymbol{q}^b)$:

$$\widetilde{\mathscr{P}}_{CL}^b\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;q^b\right) \equiv \log \frac{\widetilde{E}_b\left(q^b;\boldsymbol{p}_t\right)}{\widetilde{E}_b\left(q^b;\boldsymbol{p}_{t_0}\right)} \equiv \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}^b\right), \text{ such that } q^b = E\left(U\left(\boldsymbol{q}^b\right);\boldsymbol{p}_b\right). \tag{18}$$

Homotheticity implies that the function Hicksian demand function $\Omega_i^b\left(q^b;\boldsymbol{p}_t\right)$ does not vary in real consumption $q^b$. As such, changes in prices have an identical effect on the costs of achieving different levels of real consumption. In other words, the index $\widetilde{\mathscr{P}}_{CL}^b\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;q^b\right)$ is independent of real consumption $q^b$ corresponding to the base basket of quantities under homothetic preferences.

Nonhomotheticity implies that the Hicksian demand function $\Omega_i^b\left(q^b;\boldsymbol{p}_t\right)$ varies in real consumption $q^b$. Rich and the poor consumers have different compositions of expenditure and thus experience price inflation differently. Thus, when price changes from any period $t_0$ to $t$ are heterogeneous across goods, the changes in the cost-of-living $\widetilde{\mathscr{P}}_{CL}^b\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;q^b\right)$ may depend on the level of real consumption $q^b$. In this case, Lemma 1 still allows us to approximate the CoL price indices for base baskets of quantities that lead to levels of real consumption close to $q_{t_0}^b$ and $q_t^b$.

comparisons over long time horizons. See Appendix A.2 for a complete discussion.

[15] Note that we have $s_{it} = \Omega_i^b\left(q_t^b;\boldsymbol{p}_t\right)$, and when $\boldsymbol{p}_t \equiv \boldsymbol{p}_b$, this function also characterizes the Marshallian demand.

But we do not know how the reduced-form indices relate to the CoL price indices for base baskets of quantities that lead to substantially higher or lower levels of real consumption.

In the presence of such nonhomotheticity, the measures of welfare growth may fundamentally depend on the choice of base prices. When the values of CoL price indices depend on the base basket of quantities, the ReC indices lead to meaures of real consumption growth that vary based on their base vector of prices. Consider the example of an improvement in the welfare of a consumer in moving from period $t_0$ to $t$, i.e., a case with $U(\boldsymbol{q}_{t_0}) < U(\boldsymbol{q}_t)$. Assume also that the change in the cost-of-living in moving between the two periods is lower for richer consumers, implying in particular that $\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_{t_0}\right) > \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_t\right)$. Based on definitions (4) and (5), it immediately follows that $\mathscr{Q}_{RC}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{p}_{t_0}\right) > \mathscr{Q}_{RC}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{p}_t\right)$, that is, the growth in real consumption is higher from the perspective of the prices in period $t_0$ rather than $t$.[16]

To see the intuition for this result, note that in the base period nominal expenditures are equal to real consumption by definition, i.e. an increase in nominal expenditure translates into a one-for-one increase in real expenditure throughout the distribution of expenditures. When inflation differs for necessities and luxuries, this uniform relationship across the distribution holds only in the base period. When we pick the initial period ($t_0$) as base, given our assumption on prices (lower inflation for richer consumers), in the final period ($t$) a one-dollar increase in nominal expenditures translates into a *larger* increase in real consumption for high-income households (who have a higher spending share on luxuries) than for low-income households (who have a higher spending shares on necessities). Conversely, with the final period as base, in the initial period the one-dollar increase in nominal expenditures corresponds to a *lower* increase in real consumption for high-income households than for low-income households. Thus, for a given observed change in aggregate nominal expenditures from the initial to the final period, the change in real consumption is larger when using $t_0$ rather than $t$ as base, i.e. $\mathscr{Q}_{RC}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{p}_{t_0}\right) > \mathscr{Q}_{RC}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{p}_t\right)$.

The next lemma lies at the core of our results and characterizes the relationship between structural indices with different base vectors of prices under nonhomotheticity. We will rely on this result in the next section to construct our strategy for approximating changes in real consumption over multiple periods with a constant base vector of prices.

**Lemma 2.** *If Assumptions 1 and 2 hold, the ReC quantity index under the vector of prices $\mathbf{p}_b$ for any base period $b$ is related to the CoL price index under the initial period vector of quantities to the first*

[16]Thus, in this case, the equivalent variation (using initial prices as base) is systematically higher than the compensating variation (using final prices as base).

*order of approximation through*

$$\log \mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0}, \boldsymbol{q}_t; \boldsymbol{p}_b\right) = \frac{1}{1+\Lambda_b\left(q^b_{t_0}; \boldsymbol{p}_t\right)} \log\left(\frac{y_t/y_{t_0}}{\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_{t_0}\right)}\right) + O\left(|t-t_0|^2\left(\Delta_p+\Delta_y\right)^2\right), \tag{19}$$

*where $q^b_{t_0}$ is the real consumption corresponding to the consumption basket $\boldsymbol{q}_{t_0}$ and where the nonhomotheticity correction function $\Lambda_b$ is defined as*

$$\Lambda_b\left(q; \boldsymbol{p}_t\right) \equiv \frac{\partial \log \widetilde{\mathscr{P}}^b_{CL}\left(\boldsymbol{p}_b, \boldsymbol{p}_t; q\right)}{\partial \log q}, \tag{20}$$

*where the CoL price index $\widetilde{\mathscr{P}}^b_{CL}$ is defined by Equation (18).*

*Proof.* See Appendix A.3. □

Lemma 2 states that, to the first order of approximation, nonhomotheticitic preferences may introduce a wedge between the ReC quantity index for an arbitrary base period $b$ and the conjugate CoL quantity index $(y_t/y_{t_0})/\mathscr{P}_{CL}$, where the CoL index is defined with the initial period $t_0$ ast base. We refer to this wedge, defined by Equation (20), as the nonhomotheticity correction function. Under homothetic preferences, the correction in Equation (20) is always zero. In the presence of nonhomotheticity, the CoL price index may vary for consumers aiming to reach different baskets of base quantities, leading to a nonzero correction function in Equation (20).

Consider the case in which the base period $b$ falls fall within the interval $[t_0, t]$. In this case, the nonhomotheticity correction function is zero to the first order of approximation in $|t-t_0|(\Delta_p+\Delta_y)$ and the ReC quantity index under the base period $b$ is the dual of the CoL price index under the initial base period $t_0$ in the sense of the weak factor reversal condition. Lemma 1 applies and we can approximate both the ReC quantity and CoL price indices.

When the base period $b$ does not fall within the interval $[t_0, t]$, the corresponding nonhomotheticity correction function may be sizable. If we know the value of the correction function, we can infer the ReC quantity index based on the CoL price index for the initial base period $\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_{t_0}\right)$, on the right hand side of Equation (19). Again, we can rely on Lemma 1 to approximate the latter using reduced-form price indices, since its base period does fall within the $[t_0, t]$ interval.

The nonhomotheticity correction summarizes all the relevant information about a base period that is outside the interval of the welfare comparison. Let us next provide some intuition as to the determinants of this correction. The following lemma links the nonhomotheticity correction function to the elasticity of this demand function with respect to real consumption.

**Lemma 3.** *The nonhomotheticity correction function defined in Equation (20) satisfies:*

$$\Lambda_b\left(q^b;\boldsymbol{p}_t\right)=\int_b^t\left[\sum_{i=1}^{I}\Omega_i^b\left(q^b;\boldsymbol{p}_\tau\right)\eta_i^b\left(q^b;\boldsymbol{p}_\tau\right)\frac{d\log p_{i,\tau}}{d\tau}\right]d\tau, \tag{21}$$

*where we have defined the "real" income elasticity of demand as:*

$$\eta_i^b\left(q;\boldsymbol{p}_t\right)\equiv\frac{\partial\log\Omega_i^b\left(q;\boldsymbol{p}_t\right)}{\partial\log q}. \tag{22}$$

*Proof.* See Appendix A.3. □

Equation (21) shows that the first-order nonhomotheticity correlation function at time $t$ captures the cumulative (expenditure-weighted) *covariance between real income elasticities and price inflation* across goods, integrated from the base-year to time $t$. First, under homotheticity, the real income elasticities in Equation (22) are zero, leading to the expected result that $\Lambda_b\left(q;\boldsymbol{p}_t\right)\equiv 0$. Next, consider the case in which the price inflations are uniform across a goods, i.e., $d\log p_{i,t}/dt\equiv\gamma_t$. Even under nonhomotheticity, by definition (22) the (expenditure-weighted) mean of real income elasticities is always zero, $\sum_i\Omega_i\eta_i=0$. Thus, the correction is zero in this case as well, $\Lambda_b\left(q;\boldsymbol{p}_t\right)\equiv 0$. Since the correction is zero in these two polar cases, the effect of nonhomotheticity exists to the extent that there is a systematic and persistent relationship between real income elasticities and price inflations.

# 3 Approximating the Nonhomotheticity Correction with Income Heterogeneity

We next characterize approximations for the nonhomotheticity correction function that rely on observed variations in reduced-form price indices in a cross-section of consumers. With nonhomothetic preferences, we show how to use the cross-sectional variations in the reduced-form price indices across consumers to provide more accurate approximations to the ReC quantity indices over long horizons. We conclude this section with a simulation highlighting that our methodology accurately recovers the evolution of the exact index, without the knowledge of the parameters of the demand system.

## 3.1 Setting for the Approximation

Consider a setting where we observe the composition of consumption expenditures for a collection of consumers $n\in\mathcal{N}$ with $N\equiv|\mathcal{N}|$ heterogeneous levels of income over time. We assume that the consumers face the same sequence of prices $\boldsymbol{p}_{(0,T)} \equiv (\boldsymbol{p}_t)_{t=0}^{T}$ and the consumption patterns of consumer $n$ is characterized by the sequence of vector of quantities $\boldsymbol{q}^n_{(0,T)} \equiv (\boldsymbol{q}_t)_{t=0}^{T}$. Correspondingly, we let $y^n_t$ and $s^n_{i,t} \equiv p_{i,t} q^n_{i,t} / y^n_t$ denote the total expenditure and the share of good $i$ in total expenditure for consumer $n$ at time $t$, respectively. We assume that Assumption 1 holds with identical preferences for all consumers and that Assumption 2 holds for all consumers with the same bound $\Delta_y$ on their growth (or degrowth) in total expenditure across periods.

Consider in addition a given base period $b$ for our ReC and CoL indices. As we discussed above, the ReC quantity index allows us to consistently compare welfare, not only for a specific consumer over time, but also across all consumers. Let the sequences of utilities and real consumptions be denoted by $\left(u^n_{(0,T)}, q^{b,n}_{(0,T)}\right)_{n\in\mathcal{N}}$, where $q^{b,n}_t \equiv G_b\left(u^n_t\right)$ defines the real consumption of the consumer for base period $b$ from Equation (6).

The key insight behind the results that follow stems in the observation that nonhomotheticity correction function $\Lambda_b$ defined in Equation (20) characterizes the cross-sectional elasticity of the CoL price index with respect to real consumption $q$ under the prices in base period $b$. Having access to a cross-section of consumers that vary in income in the base period allows us to approximate this relationship using the reduced-form CoL indices, since in this period real consumption is by definition the same as the observed expenditure. This allows us to build a nonparametric approximation of the structural CoL index $\widetilde{\mathscr{P}}^{b}_{CL}(\boldsymbol{p}_b, \boldsymbol{p}_{b\pm 1}; q)$ as a function of real consumption $q$. Accordingly, we can apply the expression in Equation (19) to find the ReC index for all consumers one period away from the base period. We can then use the chaining principle to approximate the integral in Equation (21), moving from any given base period to any other period of interest.

In line with all our previous results, we provide two different approximations for the nonhomotheticity correction function, first and second order in annual inflation/expenditure growth. Despite its larger approximation error, the first-order approximation has the advantage that it requires access to a cross-section of consumers only in the base year. In contrast, the second-order approximation requires access to a cross-section of consumers overlapping every consecutive periods of interest.

## 3.2 A First-Order Nonhomotheticity Correction

We first impose the following distributional assumption on the underlying distribution of welfare in the sample of consumers.[17]

[17] In Appendix A.3, we offer an alternative set of assumptions that do not impose probabilistic restrictions on the sample of consumers.

**Assumption 3.** *For all $t \geq 0$, the real consumption across consumers has a probability distribution function that is bounded away from zero over an interval $\left[\underline{q}, \overline{q}\right]$.*

We can now introduce an algorithm that starts from the base year $b$ and computes the real consumption index period-by-period for all future periods. The core step involves a nonparametric series-function approximation, in which we fit the cross-sectional variations in the observed reduced-form geometric price index across consumers in each period to a sequence of log-power functions $\{g_k(q) \equiv (\log q)^k\}_{k=0}^{K_N}$.[18] This nonparametric functional approximation allows us to construct an approximation $\widehat{\mathscr{P}}_{b,t}^{(1)}(q) \approx \widetilde{\mathscr{P}}_{CL}^{b}(\boldsymbol{p}_b, \boldsymbol{p}_t; q)$ for the CoL index in Equation (20) and a corresponding approximation $\widehat{\Lambda}_{b,t}^{(1)}(q) \approx \Lambda_b(q; \boldsymbol{p}_t)$ for the nonhomotheticity correction function.

**Algorithm 1.** *Let $\widehat{q}_b^{b,n} \equiv q_b^{b,n} \equiv y_b^n$, define a function $\widehat{\mathscr{P}}_{b,b}^{(1)}(q) \equiv 1$, and consider sequence of log-power functions $\{g_k(q) \equiv (\log q)^k\}_{k=0}^{K_N}$ for some $K_N$ that depends on $N$, where $N$ is the number of consumers in the cross-section. For each $t \geq b$, apply the following steps:*

1. *Compute the next period function $\widehat{\mathscr{P}}_{b,t+1}^{(1)}(\cdot)$:*

$$\log \widehat{\mathscr{P}}_{b,t+1}^{(1)}(q) \equiv \log \widehat{\mathscr{P}}_{t}^{(1)}(q) + \sum_{k=0}^{K} \widehat{\alpha}_{k,t}\, g_k(q), \tag{23}$$

*where the coefficients $\left(\widehat{\alpha}_{k,t}\right)_{k=0}^{K_N}$ solve the following problem:*

$$\min_{\left(\alpha_{k,t}\right)_{k=0}^{K}} \sum_{n=1}^{N} \left( \log \mathbb{P}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right) - \sum_{k=0}^{K_N} \alpha_{k,t}\, g_k\left(\widehat{q}_t^{b,n}\right) \right)^2. \tag{24}$$

2. *Compute the real consumption in the next period for each consumer:*

$$\log \widehat{q}_{t+1}^{b,n} = \log \widehat{q}_t^{b,n} + \frac{1}{1 + \widehat{\Lambda}_{b,t+1}^{(1)}\left(\widehat{q}_t^{b,n}\right)} \log \left( \frac{y_{t+1}^n / y_t^n}{\mathbb{P}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right)} \right), \tag{25}$$

*where we have defined the approximate nonhomotheticity correction function as:*

$$\widehat{\Lambda}_{b,t+1}^{(1)}(q) \equiv \frac{d}{d \log q} \log \widehat{\mathscr{P}}_{b,t+1}^{(1)}(q) = \sum_{k=0}^{K_N} \left( \sum_{\tau=b+1}^{t+1} \widehat{\alpha}_{k,\tau} \right) \frac{d\, g_k(q)}{d \log q}. \tag{26}$$

[18] One can apply alternative series-function approximations, using alternative basis functions such as Fourier, Spline, or Wavelets. The results here generalize to such alternative nonparametric methods subject to modified regularity assumptions on the expenditure function and the distribution of real consumption across consumers (Newey, 1997).

The following proposition establishes that Algorithm 1 indeed provides an approximation to the ReC index for all $t \geq b$ as $K_N$ and $N$ go to infinity, under appropriate regularity assumptions.

**Proposition 1.** *If Assumptions 2 and 3 hold, and if the expenditure function* $\log \widetilde{E}_b(\cdot;\cdot)$ *is continuously differentiable of order* $m \geq 5$*, then as* $N$ *and* $K_N$ *grow toward infinity, the sequences of real consumptions* $\boldsymbol{q}^n_{(b,T)}$ *constructed by Algorithm 1 satisfy:*

$$\log \mathcal{Q}_{RC}\left(\boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}; \boldsymbol{p}_b\right) = \log\left(\frac{\widehat{q}^{b,n}_{t+1}}{\widehat{q}^{b,n}_t}\right) + O\left(\Delta_p^2\right) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}} \cdot \Delta_p^4 + K_N^{1-m}\right)\left(\Delta_p + \Delta_q\right)\right). \tag{27}$$

*Proof.* See Appendix A.3. □

Proposition 1 shows three sources of approximation error in the results produced by Algorithm 1: 1) the original Taylor-series approximation error in the reduced-form price index, which is second-order in $\Delta_p$, 2) the error due to the approximation of the function $\widetilde{\mathscr{P}}^b_{CL}(\boldsymbol{p}_b, \boldsymbol{p}_t; q)$ for the CoL index in Equation (20) based on the cross-section of consumers, which falls as we observe more consumers $N$, and we choose $K_N$ such that $K_N^7/N \to 0$, and 3) the error due to functional approximation using a finite set of basis functions, which falls as we choose a more flexible set of basis functions by increasing $K_N$ and thus reduce the term $K_N^{4-m}$.

Let us now examine the magnitude of the correction. The function $\widehat{\mathscr{P}}^{(1)}_{b,t+1}(q)$ provides an approximation to the CoL index from the base year to year $t+1$ for given level of real consumption $q$. This function grows roughly at the same rate as $(t+1-b) \times \Delta_p$. Since the correction in Equation (26) is also of the same order, we find that the correction becomes sizable as we deviate from the base year and $(t+1-b) \times \Delta_p$ grows. Combining this result with the insights that we gained from Lemma 3, it follows that *the size of our correction grows roughly as the product of* $(t-b)$ *and the covariance of real income elasticities and price inflations.* If the covariance is roughly of the same order as the average price inflation, e.g., around 1%, then moving a decade away from the base year leads to a correction of around 10% to the standard approximation of the year-to-year ReC index based on reduced-form indices.

**Bias in Conventional Growth Measurements** What is the implication for the bias in conventional measurements of welfare change? Let $\mathbb{Q}_G \equiv (y_{t+1}/y_t)/\mathbb{P}_G$ stand for the conjugate quantity index for the Törqvist price index. We can express the bias in this measure of real consumption growth relative to the corrected measure found based on Algorithm 1 as

$$\log\left(\frac{\mathbb{Q}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}\right)}{\widehat{q}^{b,n}_{t+1}/\widehat{q}^{b,n}_t}\right) = \lambda^n_{b,t} \times \log \mathbb{Q}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}\right), \tag{28}$$

where the value of the annual growth bias $\lambda_{b,t}^{n}$ is a monotonic transformation of the nonhomotheticity correction function:

$$\lambda_{b,t}^{n} \equiv \frac{\widehat{\Lambda}_{b,t+1}^{(1)}(\widehat{q}_t^n)}{1+\widehat{\Lambda}_{b,t+1}^{(1)}(\widehat{q}_t^n)}. \tag{29}$$

Equation (28) shows the overall size of the bias depends on two distinct forces: the size of the nonhomotheticity correction function and the size of the measured growth in real consumption. Thus, the bias is likely to be large in environments with a large covariance between price inflation and income elasticities and with fast growth in real income.

How does this bias accumulate when measuring growth over longer horizons from some period $t_0$ to $t$ using the chain rule? We can define a cumulative weighted-sum $\lambda_{b,(t_0,t)}^{c,n}$ of the annual corrected bias according to:

$$\lambda_{b,(t_0,t)}^{c,n} \equiv \sum_{\tau=t_0}^{t-1} \gamma_{\tau,(t_0,t)}^{n} \cdot \lambda_{b,\tau}^{n}, \tag{30}$$

where the weights $\gamma_{\tau,(t_0,t)}^{n}$ denotes the ratio of the uncorrected real consumption growth measured from time $t_0$ to time $t$ that is achieved between years $\tau$ and $\tau+1$:

$$\gamma_{\tau,(t_0,t)}^{n} \equiv \frac{\log \mathbb{Q}_G\left(\boldsymbol{p}_\tau, \boldsymbol{p}_{\tau+1}; \boldsymbol{q}_\tau^n, \boldsymbol{q}_{\tau+1}^n\right)}{\log \mathbb{Q}_{\mathbb{G}}{}^{c}\left(\boldsymbol{p}_{(t_0,t)}; \boldsymbol{q}_{(t_0,t)}\right)}. \tag{31}$$

Using these definitions, we can write the bias in the conventional, uncorrected, chained measures of real consumption as a share of measured growth as

$$\log\left(\frac{\mathbb{Q}_G^c\left(\boldsymbol{p}_{(t_0,t)}; \boldsymbol{q}_{(t_0,t)}\right)}{\widehat{q}_t^{b,n}/\widehat{q}_{t_0}^{b,n}}\right) = \lambda_{b,(t_0,t)}^{c,n} \times \log \mathbb{Q}_G^c\left(\boldsymbol{p}_{(t_0,t)}; \boldsymbol{q}_{(t_0,t)}\right), \tag{32}$$

where the cumulative growth bias $\lambda_{b,(t_0,t)}^{c,n}$ is given by Equation (30). As this definition makes clear, the overall bias is higher to the extent that over time periods of fast growth coincide more closely with periods of high covariance between inflation and income elasticities.

Algorithm 1 provides a simple and intuitive approach to correcting for the effects of nonhomotheticity in inferring structural price indices based on reduced-form price indices. However, our approximation error bounds here are weaker compared to the baseline case under homotheticity in Lemma 1. In the next section, we offer an extended iterative method that achieves the same error bound as that under the baseline homothetic case.

### 3.3 A Second-Order Nonhomotheticity Correction

Let us begin by revisiting Lemma 2 to present a more precise approximation to the ReC index up to the second order in the changes in prices and real consumption.

**Lemma 4.** *If Assumption 2 holds at some period $t$ and if the expenditure function $\log \widetilde{E}_b(\cdot;\cdot)$ is continuously differentiable of order at least 3, the ReC index under the base vector of prices $\boldsymbol{p}_b$ for each consumer is related to the reduced-form Törqvist price index of that consumer to the second order of approximation through*

$$\log \mathcal{Q}_{RC}\left(\boldsymbol{q}_t^{b,n}, \boldsymbol{q}_{t+1}^{b,n}; \boldsymbol{p}_b\right) = \frac{1}{1+\frac{1}{2}\left[\Lambda_b\left(q_t^{b,n}; \boldsymbol{p}_t\right) + \Lambda_b\left(q_{t+1}^{b,n}; \boldsymbol{p}_{t+1}\right)\right]} \log\left(\frac{y_{t+1}^n / y_t^n}{\mathbb{P}_T\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right)}\right) \tag{33}$$

$$+ O\left(\left(\Delta_p + \Delta_y\right)^3\right),$$

*where $\Lambda_b\left(q_t^n; \boldsymbol{p}_t\right)$ is defined by Equation (20).*

*Proof.* See Appendix A.3. □

First, note that Lemma 4 directly establishes the link between the ReC index and the reduced-form Törqvist price index, rather than the structural CoL index. More importantly, Lemma 4 introduces a second order correction for nonhomotheticity that is in turn determined by the first-order nonhomotheticity correction function defined by Equation (20). However, in contrast to the correction proposed in Lemma 2, the correction in Equation (33) also depends on the ReC index since $q_{t+1}^{b,n} = q_t^{b,n} + \mathcal{Q}_{RC}\left(\boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n; \boldsymbol{p}_b\right)$. So, even if we know the first-order nonhomotheticity correction function, the second-order correction in Equation (33) requires solving a fixed point problem to determine the ReC index.

Next, we provide an algorithm that solves the fixed-point problem corresponding to the approximation in Lemma 4 and a second-order approximation for the nonhomotheticity correction function $\Lambda_b\left(\cdot; \boldsymbol{p}_t\right)$. Algorithm 1 uses a first-order approximation of the CoL index to approximate the latter, relying on the reduced-form geometric price index. Algorithm 2 below instead builds a second-order approximation of the CoL index relying instead on the reduced-form Törqvist price index. As one may anticipate, approximating the CoL index also requires us to solve a fixed point problem, just like the case of the ReC index as laid out in Lemma 4. Thus, Algorithm 2 has an iterative structure, whereby we iteratively use Lemma 4 and the second-order approximations of the CoL index based on our current guess about the next period real consumption to update the latter.

**Algorithm 2.** *Let* $\widehat{q}_b^{b,n} \equiv q_b^{b,n} \equiv y_b^n$ , *define function* $\widehat{\mathscr{P}}_{b,b}^{(2)}(q) \equiv 1$, *and consider sequence of log-power functions* $\{g_k(q) \equiv (\log q)^k\}_{k=0}^{K_N}$ *where* $K_N$ *grows with* $N$, *the number of consumers in the cross-section. For each* $t \geq b$, *apply the following steps:*

1. *Initialize the values of the real consumption* $\widehat{q}_{t+1}^{b,n,(0)}$ *for each consumer at* $t+1$ *using Equations (23)–(26) as in Algorithm 1.*

2. *For each* $t \geq b$, *iterate over the following steps over* $\tau \in \{0, 1, \cdots\}$ *until convergence for some tolerance* $\epsilon \ll 1$*:*

   (a) *Solve for the coefficients* $\left(\widehat{\alpha}_{k,t}^{\dagger}\right)_{k=0}^{K}$ *in the following problem:*

$$\min_{\left(\alpha_{k,t}^{\dagger}\right)_{k=0}^{K}} \sum_{n=1}^{N} \left( \log \mathbb{P}_G \left( \boldsymbol{p}_{t+1}, \boldsymbol{p}_t; \boldsymbol{q}_{t+1}^n, \boldsymbol{q}_t^n \right) - \sum_{k=0}^{K} \alpha_{k,t}^{\dagger} g_k \left( \widehat{q}_{t+1}^{b,n,(\tau)} \right) \right)^2 . \tag{34}$$

   (b) *Update the next period function* $\widehat{\mathscr{P}}_{b,t+1}^{(2)}(\cdot)$*:*

$$\log \widehat{\mathscr{P}}_{b,t+1}^{(2)}(q) \equiv \log \widehat{\mathscr{P}}_{b,t}^{(2)}(q) + \sum_{k=0}^{K} \widehat{\beta}_{k,t}\, g_k(q), \tag{35}$$

   *where the coefficients* $\left(\widehat{\beta}_{k,t}\right)_{k=0}^{K}$ *solve the following problem:*

$$\min_{\left(\widehat{\alpha}_{k,t}\right)_{k=0}^{K}} \sum_{n=1}^{N} \left( \log \mathbb{P}_T \left( \boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n \right) + \lambda_t^{n,(\tau)} - \sum_{k=0}^{K} \widehat{\beta}_{k,t}\, g_k \left( \widehat{q}_t^{b,n} \right) \right)^2 , \tag{36}$$

   *with* $\lambda_t^{n,(\tau)}$ *is defined as:*

$$\lambda_t^{n,(\tau)} \equiv \frac{1}{4} \sum_{k=0}^{K} \widehat{\alpha}_{k,t}^{\dagger} \left[ \frac{d\, g_k\left(\widehat{q}_t^{b,n}\right)}{d \log q} + \frac{d\, g_k\left(\widehat{q}_{t+1}^{b,n,(\tau)}\right)}{d \log q} \right] \log \left( \frac{\widehat{q}_{t+1}^{b,n,(\tau)}}{\widehat{q}_t^{b,n}} \right). \tag{37}$$

   (c) *Update the real consumption in the next period for each consumer:*

$$\log \widehat{q}_{t+1}^{b,n,(\tau+1)} = \log \widehat{q}_t^{b,n,(\tau)} + \frac{1}{1 + \frac{1}{2}\left[ \widehat{\Lambda}_t^{(2)}\left(\widehat{q}_t^{b,n}\right) + \widehat{\Lambda}_{t+1}^{(2)}\left(\widehat{q}_{t+1}^{b,n,(\tau)}\right) \right]} \log \left( \frac{y_{t+1}^n / y_t^n}{\mathbb{P}_T\left( \boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n \right)} \right), \tag{38}$$

*where we have defined the approximate nonhomothetic correction function as:*

$$\widehat{\Lambda}^{(2)}_{b,t+1}(q) \equiv \frac{d}{d\log q}\log \widehat{\mathscr{P}}^{(2)}_{b,t+1}(q) = \sum_{k=0}^{K}\left(\sum_{\tau=b+1}^{t+1}\widehat{\beta}_{k,\tau}\right)\frac{d\, g_k(q)}{d\log q}. \tag{39}$$

(d) *Stop if* $\max_n \left|\widehat{q}^{b,n,(\tau+1)}_{t+1} - \widehat{q}^{b,n,(\tau)}_{t+1}\right| < \epsilon$ *and set* $\widehat{q}^{b,n}_{t+1} \equiv \widehat{q}^{b,n,(\tau+1)}_{t+1}$.

Function $\widehat{\mathscr{P}}^{(2)}_{b,t+1}(q)$ defined in Equation (35) provides a second-order approximation for the CoL index function $\widetilde{\mathscr{P}}^{b}_{CL}(\boldsymbol{p}_b, \boldsymbol{p}_{t+1}; q)$ defined in Equation (20). Equation (38) then relies on the results of Lemma 4 to update our current guess $\widehat{q}^{b,n,(\tau)}_{t+1}$ about the next-period real consumption. The following proposition then establishes that this iterative process yields a second-order approximation to the ReC index between any periods $t$ and $t+1$.

**Proposition 2.** *Assume that the expenditure function* $\log \widetilde{E}_b(\cdot;\cdot)$ *is continuously differentiable of order* $m \geq 5$. *If Assumptions 2 and 3 hold, then the sequences of real consumptions* $\boldsymbol{q}^{b,n}_{(b,T)}$ *constructed by Algorithm 2 satisfy:*

$$\log \mathscr{Q}_{RC}\left(\boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}; \boldsymbol{p}_b\right) = \log\left(\frac{\widehat{q}^{b,n}_{t+1}}{\widehat{q}^{b,n}_{t}}\right) + O\left(\left(\Delta_p + \Delta_y\right)^3 + \epsilon\right) \tag{40}$$
$$+ O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}}\left(\Delta^3 + K_N^{4-m}\right)^2 + K_N^{1-m}\right)\left(\Delta_p + \Delta_y\right)\right).$$

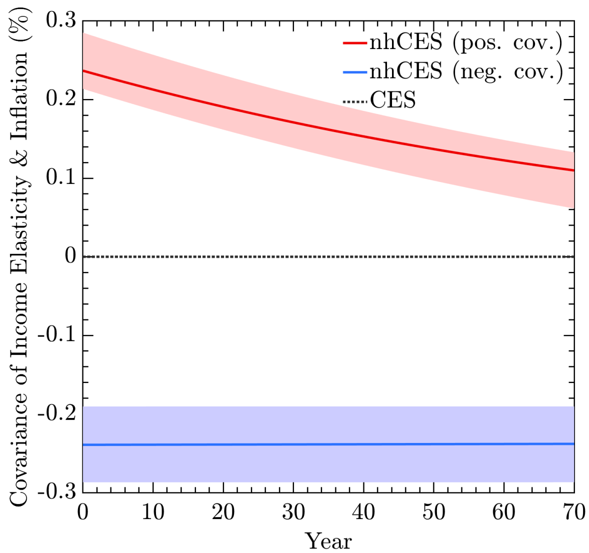
*Proof.* See Appendix A.3. □

Proposition 2 offers a substantial generalization of the index number theory to the cases involving nonhomotheticity. It offers a methodology to approximate structural price indices solely based on the cross-sectional variations in the observed reduced-form price indices across consumers, that holds across a wide class of underlying preferences.

## 3.4 Illustrative Example

To illustrate the affect of nonhomotheticity on the relationship between structural and reduced-form price indices and the performance of our algorithms over long time horizons, let us consider the following simple quantitative example. Comin et al. (2021) have shown that the nonhomothetic CES (nhCES) preferences provide a suitable account of the cross-sectional relationship between household income and the composition of expenditure among three main sectors of the economy: agriculture, manufacturing, and services. Following their specification, we assume

Figure 1: Example: The Evolution of the Income Elasticity-Inflation Covariance




*Note:* The figure compares the evolution of the covariance between the expenditure elasticities of sectoral consumption and price inflation based on the example in Section 3.4 for the two cases with positive and negative covariances against the case with homothetic preferences. The shaded area shows the standard deviation of the covariance around the mean in the population of 1,000 households.

that the expenditure function satisfies:

$$E(u;\boldsymbol{p}_t) \equiv \left( \sum_{i\in\{a,m,s\}} \omega_i \left(u^{\varepsilon_i} p_{i,t}\right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{41}$$

We use the same parameters as in Comin et al. (2021): $(\sigma, \varepsilon_a, \varepsilon_m, \varepsilon_s) = (0.26, 0.2, 1, 1.65)$, suggesting that services are luxuries (income elasticities exceeding unity) and agricultural goods are necessities (income elasticities lower than unity). We consider a population of a thousand households with an initial distribution of expenditure with a log-normal distribution with a mean corresponding to the average US per-capita nominal consumption expenditure of 3,138 in 1953 and a standard deviation of log expenditure of 0.5 (Battistin et al., 2009). We consider a horizon of 70 years and assume that over this horizon nominal expenditure grows at the constant rate of 4.48% per year (in line with the US data from the period 1953-2019). In each of the cases discussed below, we choose the fixed sectoral demand shifters $\omega_i$ in Equation (41) in such a way that in the first period the composition of aggregate expenditure fits the US average shares of sectoral consumption in the three sectors in 1953.[19] We compare the nonhomothetic specification above against a homothetic CES specification with $(\sigma, \varepsilon_a, \varepsilon_m, \varepsilon_s) = (0.26, 1, 1, 1)$.

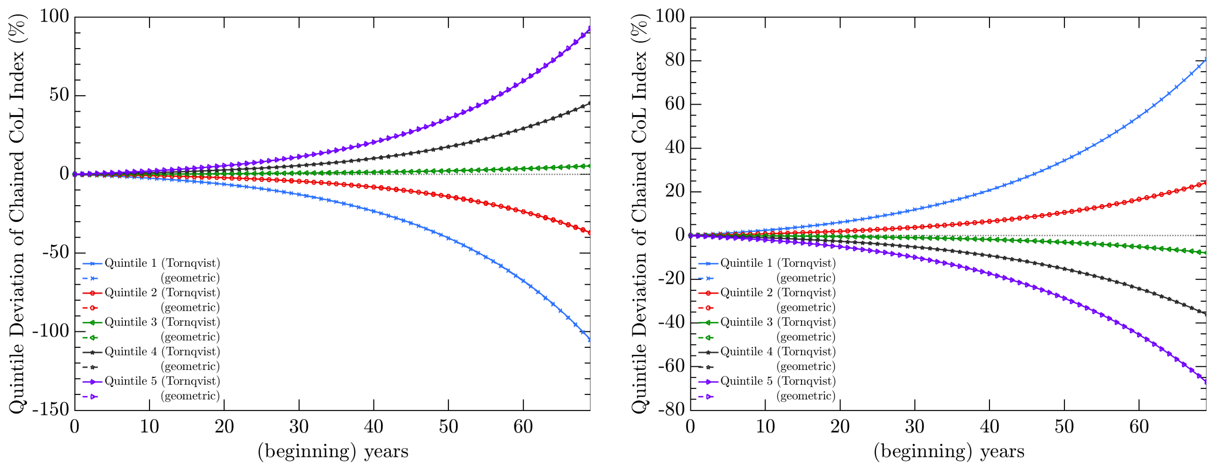
To examine the role of the covariance between price inflation and income elasticities, we apply

[19] The corresponding shares in the US based on the BLS data are 0.14, 0.27, and 0.59 for agriculture, manufacturing, and services, respectively.

Figure 2: Example: Chained Cumulative CoL Price Indices by Quintile

(a) Positive Covariance

(b) Negative Covariance



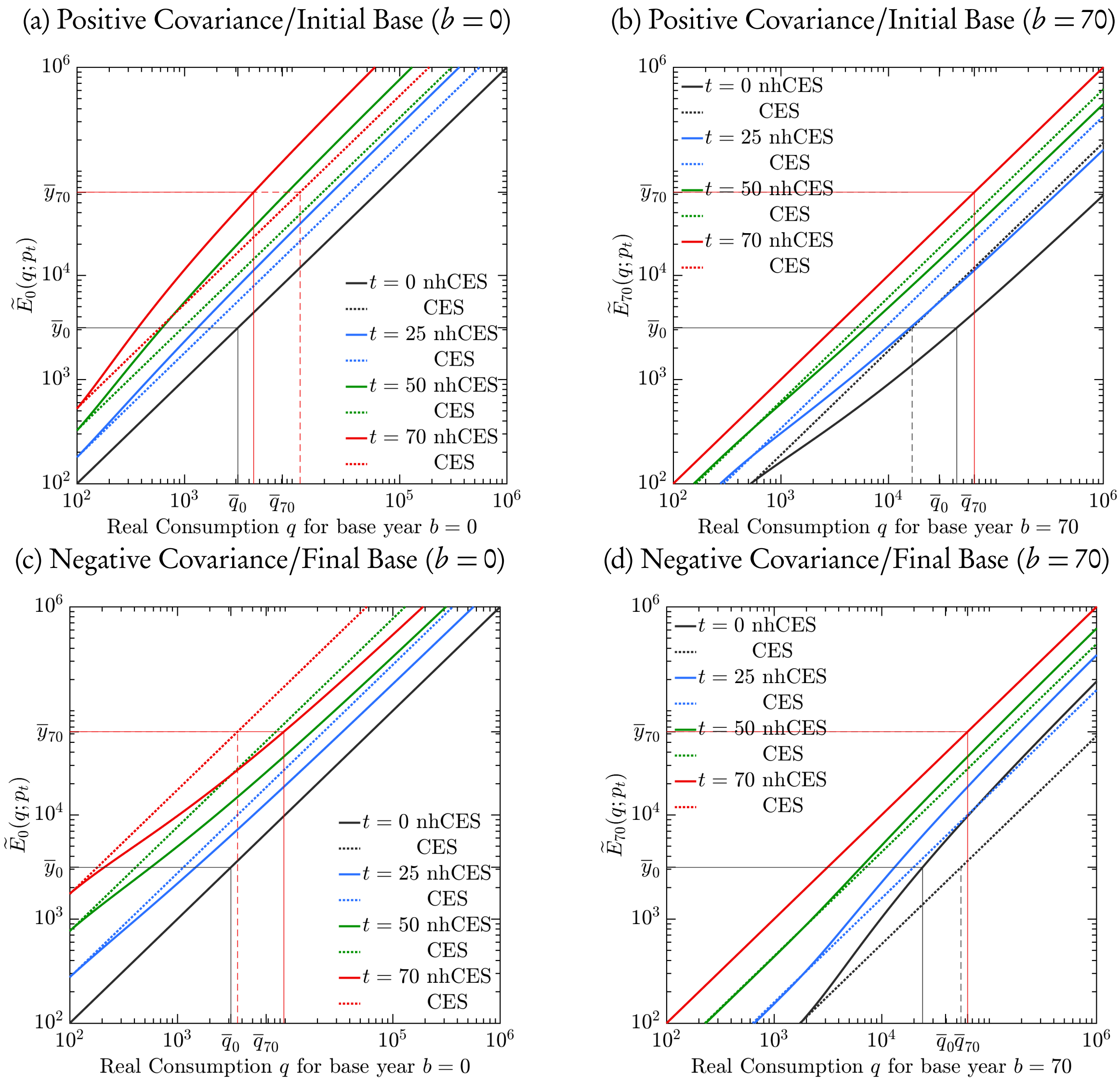

*Note:* Panels (a) and (b) show the deviation from average of the cumulative reduced-form CoL price indices for households at different quntiles of initial total expenditure over the period for the cases with positive and negative covariances of price inflations and income elasticities, respectively.

the following strategy. We set the inflation rate in manufacturing to be the average inflation rate in the US over the period 1953-2019 of 3.19%. We then consider two cases with positive and negative covariances: the inflation rates in service and agriculture are 1% higher or lower than manufacturing, respectively, in the positive case and the reverse is true in the negative case. Later in the section, we will further consider a wider range of values for these covariances.

The nonhomothetic specification implies that the income elasticity is highest for services and lowest for agriculture. Given the assumption that price inflation is highest in services and lowest in agriculture, the implication is that income elasticities positively covary with price inflation across goods. Figure 1 shows the evolution of the covariance between price inflation and the elasticity of demand for each of three goods with respect to the total expenditure of the household, where each good is weighted by its corresponding expenditure share.[20] This covariance is trivially zero in the case of homothetic preferences since the expenditure elasticities are identical and unity for all three goods. Figures 2a and 2b show how the chained reduced-form CoL price indices for households at different quintiles of initial income deviate from the population average over time. As expected, the CoL price indices are increasing in income in the case with positive covariance and decreasing in the case with negative income elasticity-covariance covariance.

[20] The expenditure elasticity of demand for good $i$ for preferences in Equation (41) is given by $\eta_{it} = \sigma + (1-\sigma)\epsilon_i/\bar{\epsilon}_t$, where $\bar{\epsilon}_t$ is the expenditure share weighted average of the $\epsilon_i$ parameters across the three goods, based on the expenditure shares in time $t$. The reason for the downward trend in the case with positive covariance is that, due to gross complementarily among the goods, the expenditure shares of households strongly shifts toward services due to both higher income elasticities and rising prices over the period. The rising concentration of consumption in services mechanically reduces the weighted covariance between price inflation and income elasticities.

Figure 3: Example: The Expenditure Function $\widetilde{E}_b(\cdot; \boldsymbol{p}_t)$




*Note:* The figure shows the expenditure function defined in terms of real consumption with the initial year as the base, for the preferences defined in Equation (41) with parameters corresponding to a nonhomothetic CES $(\sigma, \varepsilon_a, \varepsilon_m, \varepsilon_s) = (0.26, 0.2, 1, 1.65)$ (nhCES) and homothetic CES $(\sigma, \varepsilon_a, \varepsilon_m, \varepsilon_s) = (0.26, 1, 1, 1)$ functions. Panels (a) and (b) show the results for initial and final periods as the base for the case with positive income elasticity-inflation covariance, respectively. Panels (c) and (d) show the same results for the case with negative income elasticity-inflation covariance.

For the case with the positive covariance, Figures 3a and 3b compare the expenditure function, defined in terms of real consumption as in Equation (7) between the nonhomothetic and homothetic specifications, with the initial and the last periods as the base, respectively. The figures further illustrate how the corresponding expenditure functions change over time. In the homothetic case, the expenditure function always has a log-linear form (see Equation A5). Due to the overall inflation in prices, the expenditure function uniformly shifts upward over time for the homothetic CES preferences.

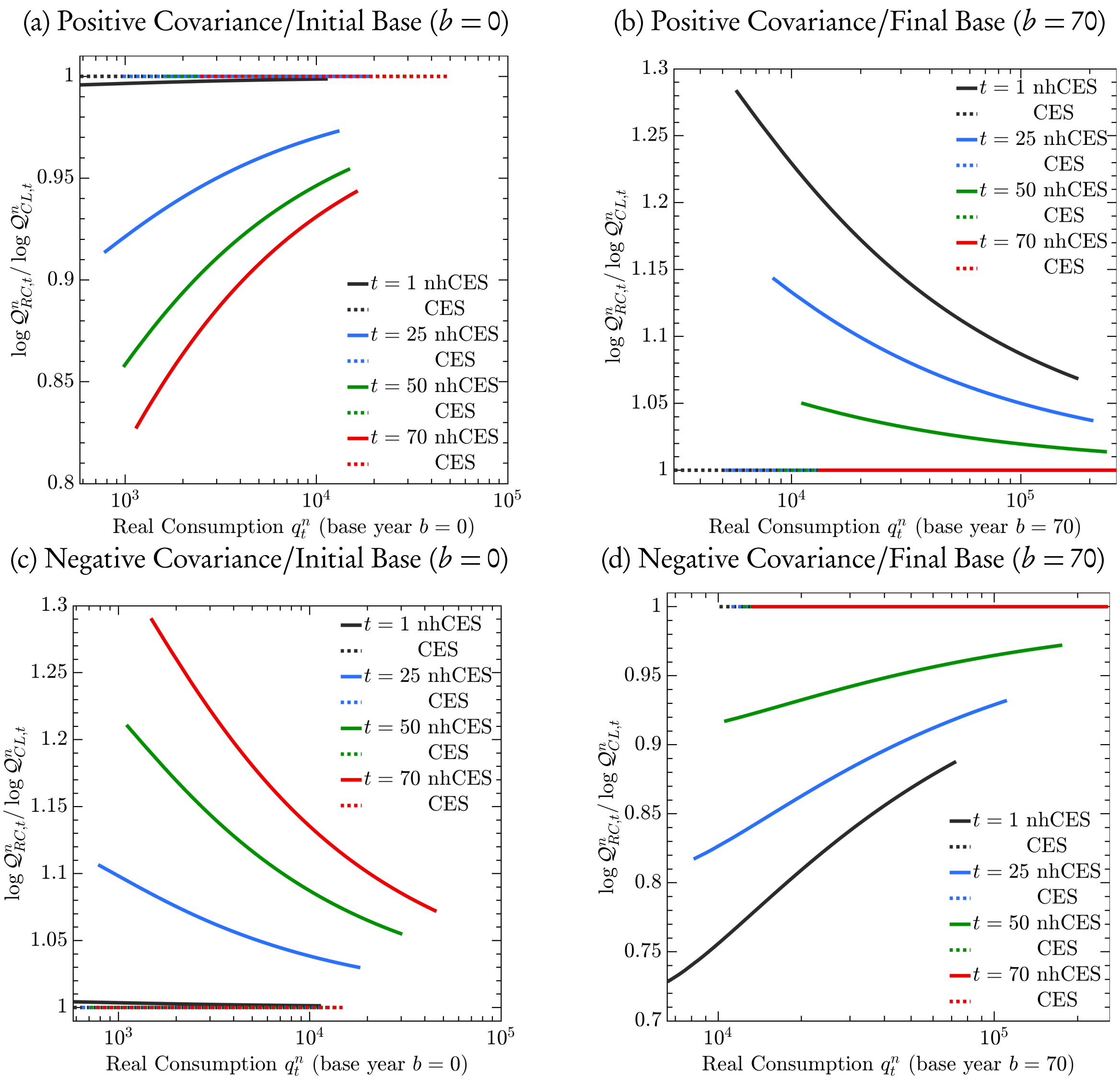
Moving to the nonhomothetic specification, first let us consider the initial period as the base as in Figure 3a. By definition, the expenditure function begins with the same log-linear form in the initial base period. As time passes, the costs of achieving higher levels of real consumption increasing rises relatively faster since to achieve these levels households want to shift their consumption toward goods with faster growing prices (higher inflation). Thus, the expenditure function defined in terms of real consumption in Equation (7) increasingly deviates from linearity and becomes more convex as time passes. As the figure shows, the upward shift in the expenditure function is larger compared to the homothetic case for higher levels of real consumption.

Next, consider the final period as base as in Figure 3b. By definition, in this case the expenditure function has a log-linear form in the final period. As we move backward in time, the costs of achieving higher levels of real consumption falls relatively faster since to achieve the corresponding levels of welfare households want to shift their consumption toward goods that are falling faster (higher inflation). Thus, the expenditure function increasingly deviates from linearity and becomes more concave as we move toward the initial period. Crucially, regardless of the choice of the base period, with nonhomothetic preferences and with a positive income elasticity-inflation covariance, the expenditure function is more convex in the later periods.

The figures further illustrate the process of the determination of the growth in real consumption, in the specific case of the observed average nominal expenditures in the US data, which rises from $\overline{y}_0 = 3,138$ to $\overline{y}_{70} = 63,036$ dollars. In Figure 3a, the real consumption is identical to the nominal expenditure in the base period ($t = 0$) for both preferences, but maps to two distinct values in the final period depending on whether the preferences are homothetic or nonhomothetic. The situation is reversed in Figure 3b, whereby the real consumption is identical to the nominal expenditure in the final period ($t = 70$) for both preferences, but maps to two distinct values in the initial period. Due to the positive covariance between inflation and income elasticities, regardless of the choice of the base period, the growth in the average real consumption is *lower* for households whose preferences are characterized by nonhomotheticity. This is despite the fact that households share the same evolution of nominal expenditures and sectoral prices under both sets of preferences.

Figures 3c and 3d examine the same patterns in the case with a negative covariance between price inflations and income elasticities. In this case, the expenditure function becomes more concave as time progresses, since now consumers shift the composition of their expenditures toward goods that have lower price inflations. With the initial period as base, the expenditure function begins with a log linear form and becomes more concave as we move forward in time. With the final period as base, the expenditure function ends with a log linear form in the last period and becomes more convex as we move backward in time. Regardless of the choice the base, the growth in average real consumption is *higher* for households whose preferences are characterized

Figure 4: Example: The Nonhomotheticity Correction

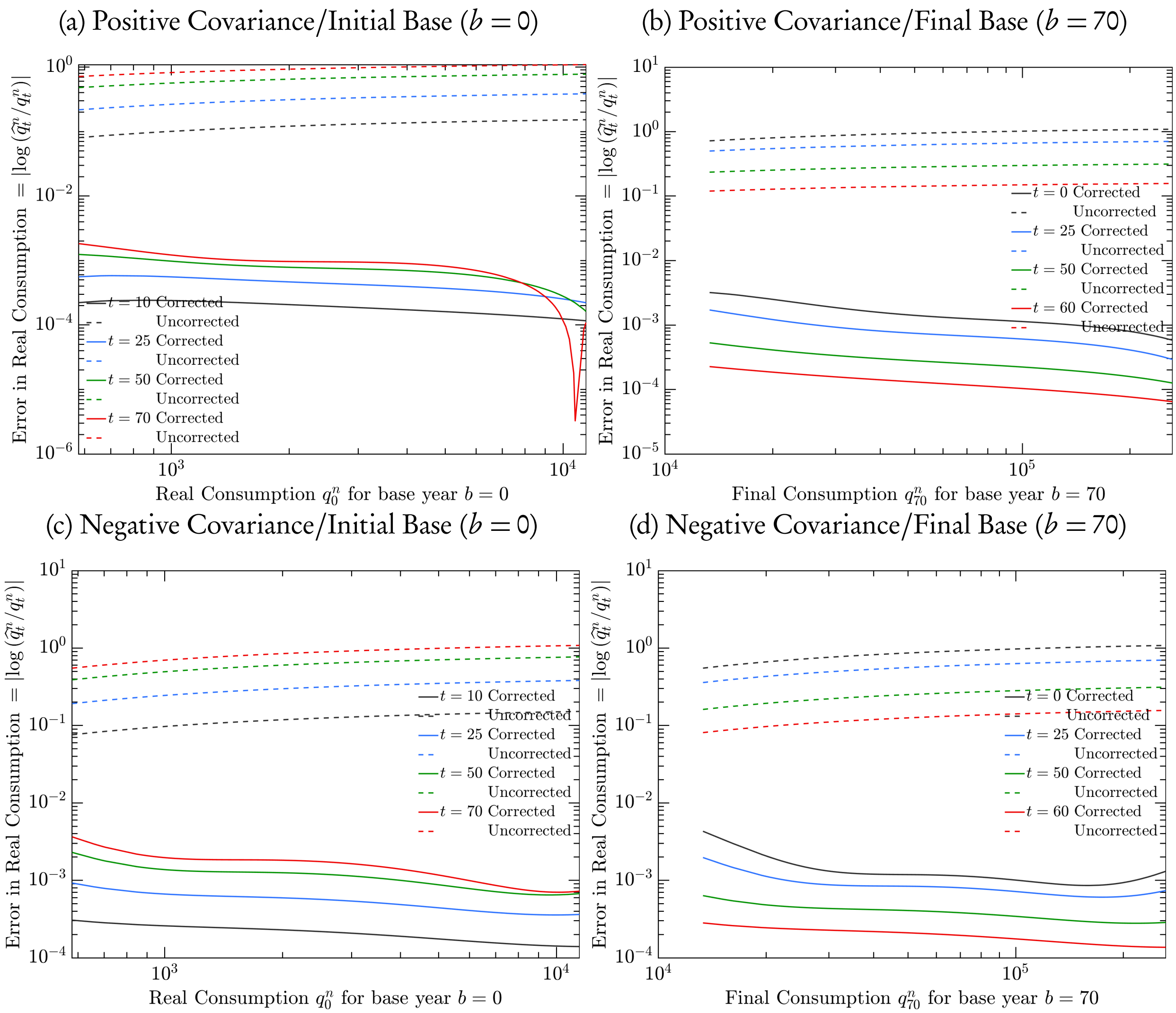



*Note:* The figure shows the ratios of the ReC and CoL quantity indices across households as a function of the current real consumption for the choices of base period (a) $b = 0$ and (b) $b = 70$ with the positive income elasticity-inflation covariance and (c) $b = 0$ and (d) $b = 70$ with the negative covariance. The underlying data is drawn based on preferences defined in Equation (41) with parameters corresponding to a nonhomothetic CES $(\sigma, \varepsilon_a, \varepsilon_m, \varepsilon_s) = (0.3, 0.2, 1, 1.7)$ (nhCES) and homothetic CES $(\sigma, \varepsilon_a, \varepsilon_m, \varepsilon_s) = (0.3, 1, 1, 1)$ functions.

by nonhomotheticity.

Figures 4a and 4b examine the implications of nonhomotheticity for the relationship between ReC and CoL indices from Lemma 2 in the case with the positive covariance. In particular, the figure for each household shows a monotonic transformation of the nonhomotheticity correction function, which is the ratio of the logarithm of the ReC index, i.e., the left hand side of Equation (19), to the logarithm of the CoL quantity index, i.e., $\log(y^n_{t+1}/y^n_t) - \log \mathscr{P}^n_{CL,t}$ on the right hand side of Equation (19). This ratio corresponds to the nonhomotheticity correction $\left(1 + \Lambda^n_{b,t}\right)^{-1}$

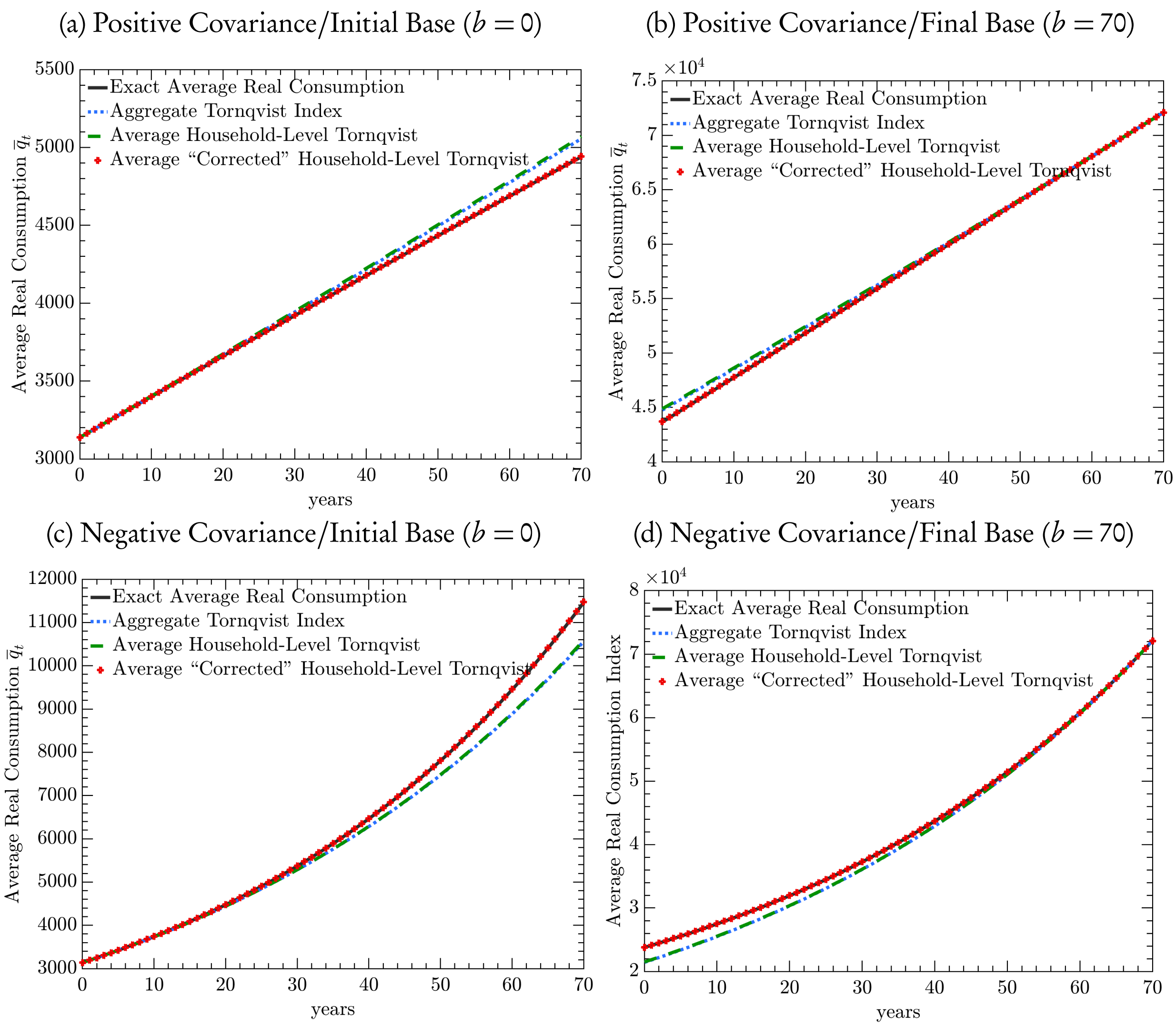
Figure 5: nhCES Example: Nonparametric Approximation of Real Consumption




*Note:* The figures compare the error in the approximate value of real consumption between the standard reduced-form geometric price index and the one corrected based on the first-order Algorithm 1. The correct value of real consumption is calculated based on the underlying parameters of the nhCES preferences. The panels show the error for the choices of base period (a) $b = 0$ and (b) $b = 70$ with the positive income elasticity-inflation covariance and (c) $b = 0$ and (d) $b = 70$ with the negative covariance.

defined in Lemma 2. This ratio is unity for all households at all times for the homothetic CES preferences, as predicted by Equation (A6). In the nonhomothetic case, the ratio is less or greater than unity depending on the choice of the base year. In particular, the CoL quantity index formula overestimates the ReC quantity index with the initial period as the base, and underestimates it with the final period as the base. The size of this bias is larger for poorer households and grows substantially for all households over time. The former result is in line with the fact that poorer household face a higher dispersion in the sectoral composition of their expenditures, leading to a higher expenditure-weighted covariance of price inflation and income elasticities (Lemma 3). Figures 4c and 4d show similar results for the case with the negative covariance, but with a reverse

Figure 6: nhCES Example: The Evolution of Average Real Consumption

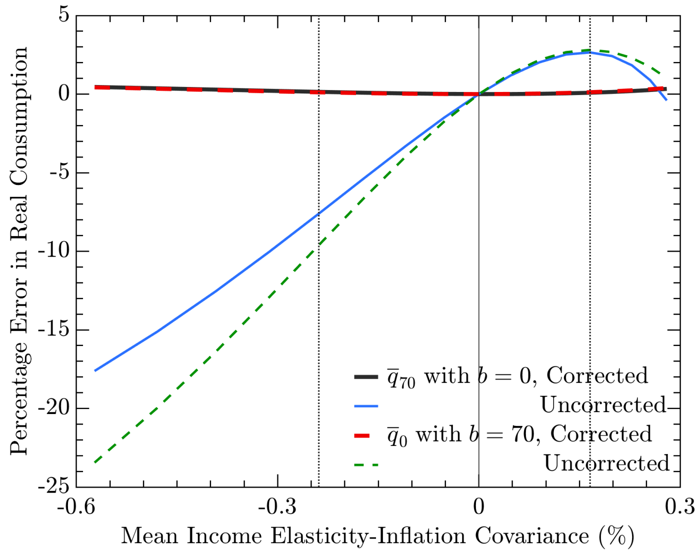



*Note:* The figures compare the evolution of the correct value of average real consumption with three different approaches to approximating this value. The standard approach relies on the reduced-form Törnqvist index defined in terms of aggregate consumption expenditure shares. The average household-level approach uses household-specific expenditure shares, and the "corrected" approach additionally uses the second-order nonhomotheticity correction. The panels show the resulting series for the choices of base period (a) $b = 0$ and (b) $b = 70$ with the positive income elasticity-inflation covariance and (c) $b = 0$ and (d) $b = 70$ with the negative covariance.

sign for the values of $\Lambda^n_{b,t}$.

Figures 5a-5d compare the error in the approximations of ReC index based on the chained reduced-form geometric price index with those found using the first-order nonhomotheticity correction following Algorithm 1 for different base periods and income elasticity-inflation covariances. Here, we use the underlying preference parameters to compute the correct value of the real consumption $q^{b,n}_t$ for each household $n$ at each point in time $t$, and compare that value with the approximate value $\widehat{q}^{b,n}_t$ found by either of the two strategies in each case. As we can see, the standard approach leads to substantially larger errors in the inferred measures of real consump-

Figure 7: Example: Real Consumption Error and Income Elasticity-Inflation Covariance




*Note:* The figure compares the error in the corrected and uncorrected approximations of the average final and initial real consumption for the initial and final periods as base, respectively, as a function of the mean covariance between price inflations and expenditure elasticities over the period.

tion. After 70 years, this error grows for some households to be of the same order of magnitude as the correct real consumption. Applying the simple first-order correction of Algorithm 1 reduces the error by several orders of magnitude. Figures A1a-A1d in Appendix C compare the sizes of the error when using the first-order approximation approach of Algorithm 1 and that found by the recursive approach of Algorithm 2 in each case, showing that the second-order approximation generally leads to lower approximation errors.

Figures 6a-6d examine the implications for aggregation across households. We compare the evolution of the correct ReC index over time with a number of different approximations. First, we see that an approach based on using the reduced-form Törnqvist index using *aggregate* sectoral composition of consumption expenditure and the one using the average household-level reduced-form Törnqvist index yield similar results. Both these approaches lead to sizable overestimation or underestimation of the growth in real consumption depending on the choice of the base period or the covariance between price inflation and income elasticities. Finally, applying our second-order nonhomotheticity correction yields results that are virtually indistinguishable from the correct evolution of ReC index. Thus, our approach accurately recovers the evolution of the exact index *without the knowledge of the parameters of the demand system.*

To show how the results extend to other ranges of the values of covariance between price inflations and expenditure elasticities, we perform one last exercise with our illustrative example. We consider alternative trends in prices varying the deviations between inflation in services and agriculture from that in manufacturing (fixed to the average level of 3.19%) symmetrically from -2%

to +2%. In each case, we apply the same analysis as before, computing the uncorrected chained reduced-form ReC indices as well as our first-order corrected approximation of the ReC index. Figure 7 shows how the error in the approximated values of average real consumption in the final period (when the initial period is taken as base) and the initial period (when the final period is taken as base) vary with the corresponding mean of the covariance between income elasticities and price inflations over the period. The figure also indicates the two cases corresponding to the the positive and negative covariance settings studied so far with dotted black lines.

As expected, when income elasticities are uncorrelated with price inflations, the uncorrected reduced-form ReC quantity indices approximate the correct values with negligible errors. However, as the covariance deviates from zero, the error grows for the uncorrected reduced-form indices. As the covariance falls to around -0.6% per year, the error in the uncorrected reduced-form indices grows to around 20% of the average real consumption across households. As the covariance grows above zero, the error initially rises. The error in the uncorrected reduced-form index ultimately begins to fall for large and positive values of covariance. This is because those scenarios lead to negligible growth in average household real consumption, which mechanically reduces the size of the bias in the reduced-form indices as we saw in Equation (28).[21] In contrast, the error in the approximation achieved with our nonhomotheticity correction remains close to zero over the entire range of values of the covariance.

# 4 Empirics

## 4.1 Data

To assess the importance of correcting for nonhomotheticity on the measurement of inequality and long-run growth in welfare, we build a dataset providing total expenditures and expenditure shares for a consistent set of products over time, covering the entire consumption basket of households in the United States. Such a dataset is not readily available from public statistics, due to two challenges. First, the product classification in surveys of consumer expenditures, such as the Consumer Expenditure Survey (CEX), is not the same as the classification used in the price surveys of the Bureau of Labor Statistics. Second, the definition of product categories changes over time in both the expenditure surveys and price surveys.

To address these challenges, we build consistent categories over time and a crosswalk delivering a linked dataset of the Consumer Expenditure Survey (CEX) to the Consumer Price Index (CPI) categories. Our preferred dataset tracks 19 categories from 1953 to 2019. Expenditure

[21]Figure A2 in Appendix C shows the overall growth in average real consumption over the period as a function of the income elasticity-inflation covariance. For positive values of the covariance, the growth diminishes toward zero.

shares by income quintiles are available each year from 1984 onward. Prior to this date, we use our first-order approximation to the correction for nonhomotheticities, using expenditure shares observed in the 1980s. In addition to implementing the correction for nonhomotheticities, this dataset allows us to compute the inequality in reduced-form price indices over a long time horizon, thus extending prior estimates that have focused on much shorter time series. Finally, to measure the growth rates of consumption by income quintiles, we use the CEX from 1984 onward. Prior to 1984, due to data limitations, for all income groups we use the aggregate growth rate of consumption expenditure per capita, as measured by the Bureau of Economic Analysis (BEA). Appendix B provides a detailed description of the data sources and crosswalks used for our main linked datasets as well as for sensitivity analysis.

To assess the sensitivity of our findings to other data construction choices, we build and study two alternative datasets. First, instead of relying only on the expenditure weights from the Consumer Expenditure Survey, we use the aggregate category-level consumption weights used by the Bureau of Labor Statistics in the official CPI index. To obtain corrected expenditure shares for each income quintile, we rescale these aggregate weights across income quintiles using CEX expenditure weight. This approach allows us to perfectly match the aggregate expenditure weights used in the official CPI index, but we can implement it only at the level of 10 broad expenditure categories from 1953 to 2020.

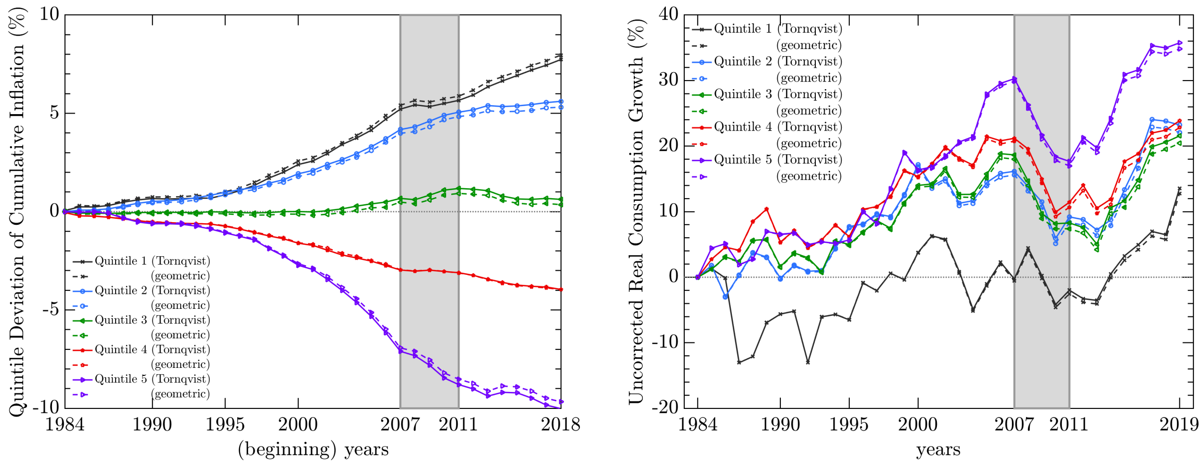
In a second robustness check, we implement our non-homothetic correction for a subset of expenditures for which product-level data is available, using Nielsen data covering consumer packaged goods, or about 15% of aggregate expenditure. This robustness check is motivated by prior work showing that most of the heterogeneity in inflation rates arises at the product level, within detailed product categories (Jaravel, 2019). We assess whether using product-level data meaningfully affects the size of the bias we estimate, at the cost of restricting attention to a subset of total expenditure. To implement this robustness check, we work with the Nielsen data from 2004 to 2015.

## 4.2 Main Results

**Measuring Growth in the Short and Medium Run** Figure 8a compares the deviation of the chained reduced-form price indices for different quintiles of income from 1984 to 2019 from the population mean over time. First, we find that reduced-form price indices are decreasing in income over the period, i.e. there is *a negative* covariance between income elasticities and price inflation across goods. Over the course of these 35 years, a gap of around 20 of percentage points has opened up in the chained reduced-form indices between the lowest and the highest quintiles of income. This finding is consistent with the growing literature on "inflation inequality," the fact

Figure 8: CEX and BLS Data: Conventional Price and Quantity Indices

(a) Quintile Deviation of Price Indices (b) Quintile Quantity Indices




*Note:* Panels (a) and (b) show the evolution of cumulative reduced-form price and quantity indices for each quintile of income over the period, respectively. The great recession has been indicated in grey background.

that inflation rates are higher for lower-income households (e.g., Kaplan and Schulhofer-Wohl, 2017; Jaravel, 2019; Argente and Lee, 2021); our data shows that this trend persists over several decades. Based on our discussions in Section 2, we should expect that reduced-form quantity indices underestimate the values of final and initial real consumption when we consider the initial and final period prices as the basis for the welfare comparisons.[22]

Figure 8b compares the evolution of the corresponding reduced-form quantity indices, capturing the growth in real consumption based on the conventional measures, across quintiles over the period. The figure clearly shows that the consumption inequality has risen for nearly three decades (1984-2014), over which real consumption stagnated for households in the lowest quintile while growing for richer households. Whereas the real consumption of households in the highest quintile has grown by around 35% over the entire period, the corresponding growth for the lowest quintile is only around 13%.

Figures 9a and 9b show the evolution of the annual bias in reduced-form indices of real consumption growth for initial (1984) and final (2019) periods as base, respectively. The sizes of these biases are expressed as a share of measured growth, as given by $\lambda^n_{b,t}$ defined in Equation (28) for each quintile $n$ at each time $t$ throughout the period.[23] Here, we have used a quadratic function

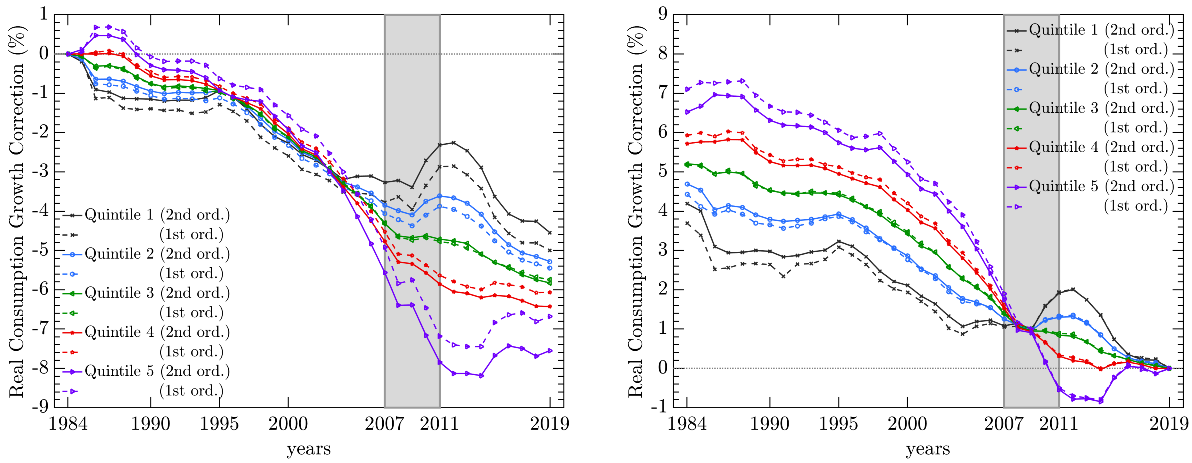
[22] Note that, although the cumulative level of inflation inequality shown in Figure 8a is economically meaningful, it is smaller than the deviations we considered in the illustrative example in Figures 2a and 2b.

[23] In the case of the second order approximation, the value of the bias is given by:

$$\lambda^n_{b,t} \equiv \frac{\frac{1}{2}\Lambda^{(2)}_t\left(\widehat{q}^n_t\right) + \frac{1}{2}\Lambda^{(2)}_{t+1}\left(\widehat{q}^n_{t+1}\right)}{1 + \frac{1}{2}\Lambda^{(2)}_t\left(\widehat{q}^n_t\right) + \frac{1}{2}\Lambda^{(2)}_{t+1}\left(\widehat{q}^n_{t+1}\right)}.$$

Figure 9: CEX and BLS Data: Bias in Reduced-form Real Consumption Growth

(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the evolution of the annual bias in the reduced-form measures of real consumption growth $\lambda^{n}_{b,t}$, defined in Equation (28), for different quintiles of income for the initial and final years as the base period.

for fitting the cross-sectional variations in price indices.[24]

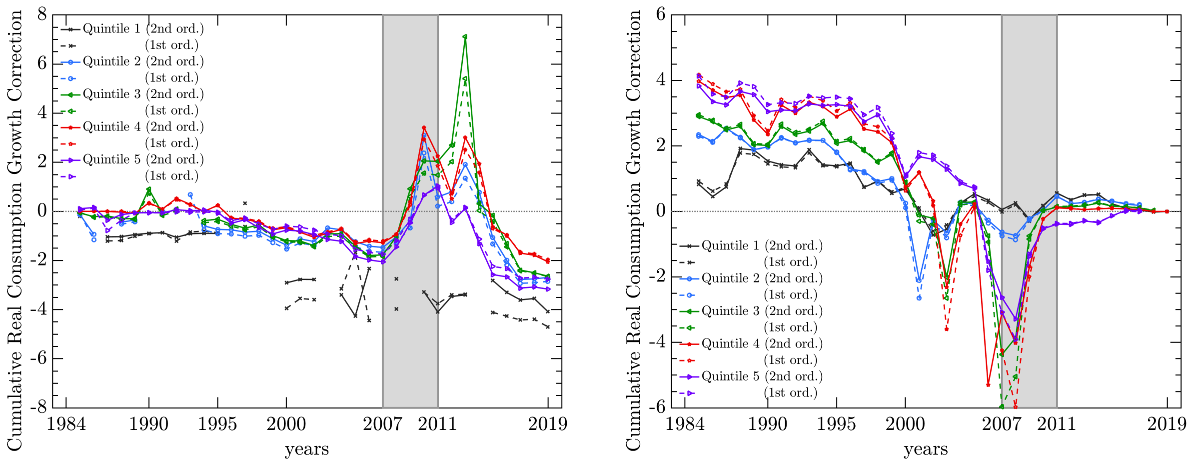
More importantly, the values of bias are negative and positive depending on whether the base in the initial or the final period, respectively. The signs of the measured biases are in line with the discussion of our illustrative example in Section 3.4, for the case with negative covariance of income elasticities and price inflations (see Figures 4c and 4d). The magnitudes of the bias (and the required correction) grow over time, particularly for households in higher income quintiles. For the highest income quintile, the uncorrected indices underestimate the annual growth in the final year and overestimate it in the initial year by around 8%, when the base period is taken to be the initial and the final periods, respectively. The sizes of the biases for the lowest quintile falls to roughly 4% of the measured growth.

Figures 10a and 10b show the corresponding values of cumulative bias (and correction) $\lambda^{c,n}_{1984,(1984,t)}$ and $\lambda^{c,n}_{2019,(t,2019)}$ for the initial and final base vectors of prices, defined in Equations (30) and (31). As we may expect, the cumulative bias grows in tandem with the annual bias over the period, albeit at a slower pace due to the fact that the former is a weighted average of the latter. The patterns become disorderly during the great recession since the weights in Equation (30) become negative due to the fall in real consumption over this period. But over the entire horizon, the

[24]Figures A3a and A3b show in Appendix C the values of the first and second order nonhomotheticity corrections $\left[1+\Lambda^{(1)}_{t+1}\left(\widehat{q}^{n}_{t}\right)\right]^{-1}$ and $\left[1+\frac{1}{2}\Lambda^{(2)}_{t}\left(\widehat{q}^{n}_{t}\right)+\frac{1}{2}\Lambda^{(2)}_{t+1}\left(\widehat{q}^{n}_{t+1}\right)\right]^{-1}$ for initial (1984) and final (2019) periods as base, respectively. Figures A4a and A4b in Appendix C show the first and second order nonhomotheticity corrections where we instead use a linear approximation for fitting the cross-sectional variations in price indices, in which case the correction is identical for all quintiles. The figure shows that the values of the bias are fairly similar for both approximations.

Figure 10: CEX and BLS Data: Cumulative Bias in Reduced-form Real Consumption Growth

(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the evolution of the cumulative bias $\lambda^{c,n}_{1984,(1984,t)}{}''^{n}$ and $\lambda^{c,n}_{2019,(t,2019)}{}''^{n}$ in the reduced-form measures of real consumption growth $\lambda^{n}_{b,t}$, defined in Equation (28), for different quintiles of income for the initial and final years as the base period. To avoid showing outlier cases, we have dropped quintile-year observations for which the overall real consumption growth, the denominator in Equation (32), is smaller than 0.01.

uncorrected measures underestimate growth by up to 4% with respect to the initial base prices, while overestimating it by up to 4% with respect to the final base prices.

Figures 11a and 11b show the evolution of the relative correction in each quintile's *level* of real consumption. These values correspond to the relative size of the corrected ReC indices compared to the uncorrected indices, when expressed in terms of the initial and final periods, respectively.[25] The figure shows the relative values of real consumption based on both the first and second order approximations. With few exceptions, the uncorrected measures *underestimate* the value of real consumption. For instance, when we express real consumption in terms of constant 1984 base prices, the magnitude of this underestimation grows exponentially over time all the way until the great recession, where it falls substantially as the real consumption of households also collapses. It then rebounds quickly over the last decade, growing to about 1 percentage point for the highest quintile.[26]

Figures 12a and 12b compare the evolution of *average* real consumption across quintiles as given by corrected and uncorrected measures, when expressed relative to the uncorrected measure of *aggregate* real consumption.[27] The former measures average real consumption across quintiles

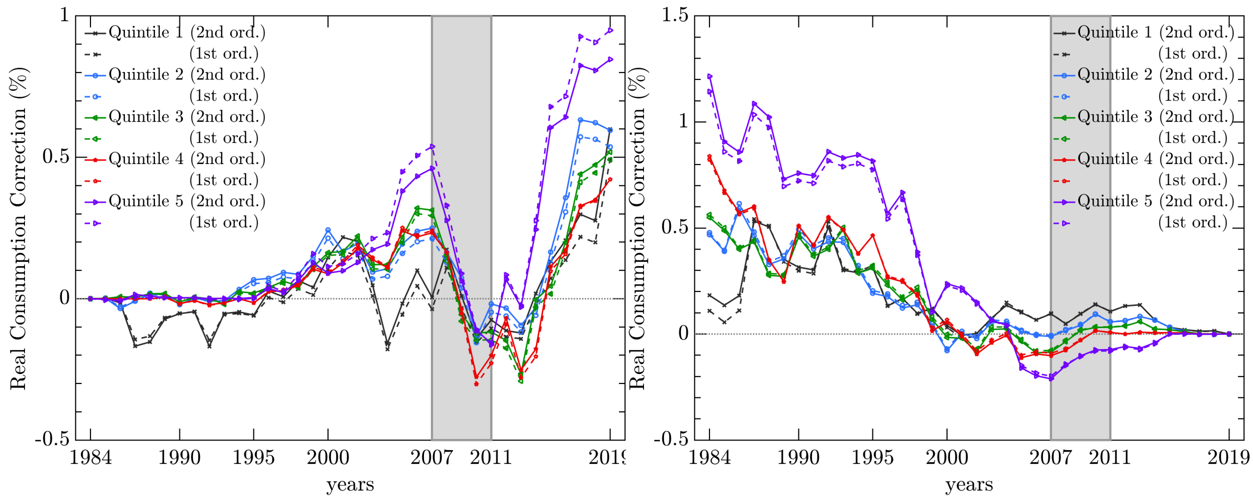
[25]Figures A5a and A5a in Appendix C show the same results when we use a linear approximation for fitting the cross-sectional variations in price indices.

[26]The contribution of the correction to the measurement of inequality in real consumption is negligible in the case with 1984 prices as base. With 2019 prices as base, the correction lowers the measured rise in inequality of real consumption by around 5% (1 percentage points lower from a baseline of 19.5 percentage points).

[27]Figures A6a and A6b in Appendix C presents the same graph using a linear, as opposed to a quadratic, approximation for fitting the cross-sectional variations in price indices

Figure 11: CEX and BLS Data: Corrected relative to Uncorrected Real Consumption

(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the ratio of the corrected to the uncorrected measures of real consumption vary over time for each quintile of initial real consumption for the initial and final years as the base period, respectively. The great recession has been indicated in grey background.

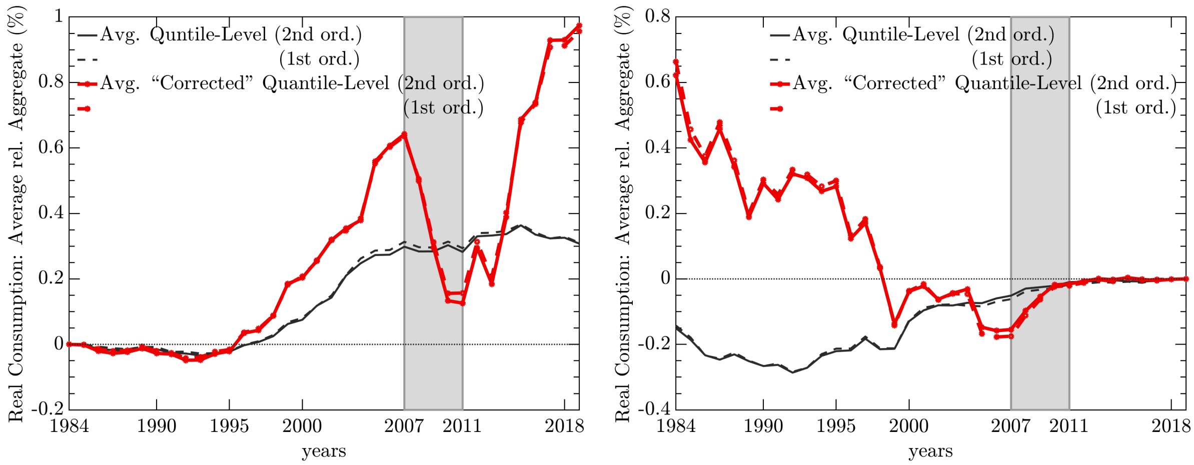
of income based on each the composition of expenditure in each quintile. The latter measure is real consumption based on the aggregate composition of consumption expenditure. Relative to the aggregate measure that ignores heterogeneity, the average real consumption found using quintile-level, uncorrected indices imply higher growth over the period, i.e., higher final real consumption with initial base prices and lower initial real consumption with final base prices. This pattern is driven by the fact that households at higher quintiles have witnessed disproportionately higher growth in their real consumption. More importantly, our nonhomotheticity correction suggests that the uncorrected measures, even when accounting for heterogeneity, underestimate average real consumption. The gap between the corrected and uncorrected approximations of average real consumption grows to around 0.8% over the period.

**Measuring Growth in the Long Run** Our next exercise is to extend the horizon of our comparisons of real consumption all the way back to the early 1950s. As previously indicated, we do not have access to systematic disaggregated data on the household-level composition of consumption expenditure before 1984. To utilize our approximation of nonhomotheticity correction, we apply the following approximation to express real consumption in terms of constant 1984 prices:

$$\log \mathcal{Q}_{RC}\left(\overline{\boldsymbol{q}}_{t+1}, \overline{\boldsymbol{q}}_t; \boldsymbol{p}_{1984}\right) \approx \frac{1}{1+\widehat{\Lambda}_{1984,t}\left(\overline{q}_{t+1}\right)} \log\left(\frac{\overline{y}_t / \overline{y}_{t+1}}{\mathbb{P}_G\left(\boldsymbol{p}_{t+1}, \boldsymbol{p}_t; \overline{\boldsymbol{q}}_{t+1}, \overline{\boldsymbol{q}}_t\right)}\right), \qquad t < 1984, \quad (42)$$

Figure 12: CEX and BLS Data: The Evolution of Average Real Consumption

(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* The figure shows the evolution the average corrected and uncorrected measures of real consumption across quintiles relative to the measure of aggregate real consumption that ignores income heterogeneity. The latter defines the reduced-form index of real consumption using aggregate consumption expenditure shares. Panels (a) and (b) show the correction for the initial and final years as the base period, respectively. The reduced-form prices indices used for the 2nd and 1st order approximations are geometric and Tornqvist indices, respectively.

To find the estimated value of the nonhomotheticity correction function $\widehat{\Lambda}_{1984,t}\left(\overline{q}_{t+1}\right)$ in Equation (42), we apply Algorithm 1 over a single period between 1984 and any given year $t$, using the 1984 cross-section of households and the cumulative inflation in prices from 1984 and that year. We alternatively consider expressing real consumption in constant 2019 prices, in which case we apply Algorithm 1 to chain the observed cross sections between 2019 and every year after 1984 and then use the approximation above to extend the analysis to the years before 1984. Due to the absence of cross-sectional data in the intervening period, we use a linear approximation of the cross-sections of price indices as our baseline analysis, to avoid overfitting the composition of consumption in 1984 to the previous decades.

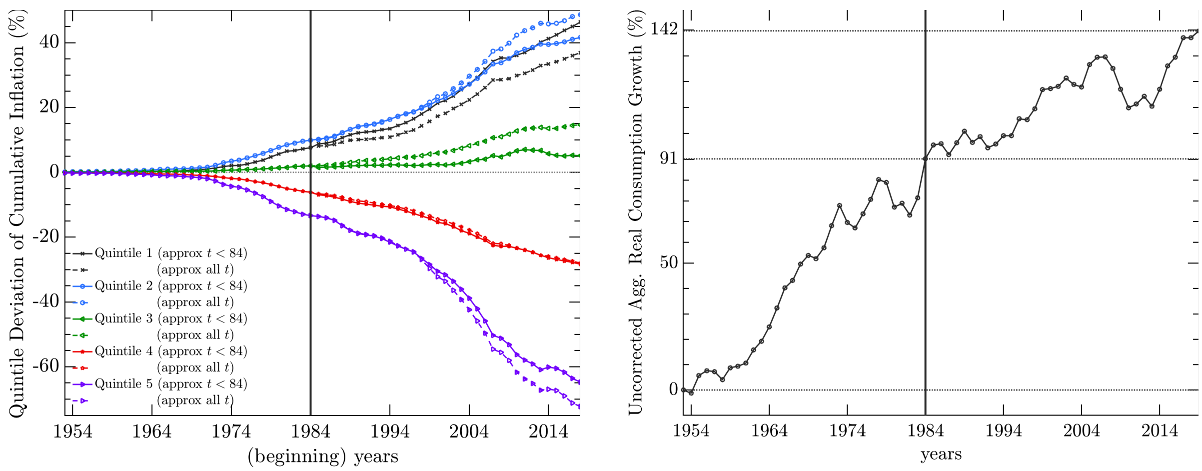
Figure 13a shows how the evolution of the gaps in the cumulative geometric price indices across different quintiles over the entire time horizon from 1954 to 2019. As explained above, in the absence of cross-sectional data before 1954, we use the 1984 cross-section to compute the index over the entire period. For the post-1984 period, the figure compares this unchained index against the chained geometric index that uses the cross-sectional data between 1984 and 2019. Over this period, we find that the two yield similar pictures for the inflation-income relationship: the overall inflation gap between the highest and the lowest quintile based on the unchained index is within 1% of the value of this gap based on the chained index.

Figure 13b shows the growth in aggregate real consumption as measured by the reduced-form quantity index, the conjugate of the geometric price index, from 1954 to 2019. The growth in the first three decades is more rapid, reaching to 91% by 1984. The growth in the remainder of

Figure 13: CEX and BLS Data: Conventional Indices Over the Extended Horizon (1954-2019)

(a) Deviation of Price Indices from Mean

(b) Aggregate Quantity Indices




*Note:* Panel (a) shows the deviation of cumulative geometric price indices for different quintiles from the mean, where we have used the 1984 cross-section to define the expenditure shares for each quintile (approx all $t$). For comparison, the figure also shows the chained indices using all the cross-sections after 1984 (approx $t < 84$). Panel (b) shows the growth in real consumption as measured using the corresponding cumulative quantity index for the aggregate composition of consumption expenditure. The growth in years 1984 and 2019 are indicated.

the period is slower, such that the cumulative growth by 2019 only reaches 142%.

Figure 14a shows the evolution of the annual $\overline{\lambda}_{b,t}$ and cumulative $\overline{\lambda}^{c}_{b,(t_0,t)}$ growth bias in reduced-form measures of aggregate real consumption growth based on our nonhomotheticity correction scheme. First, let us consider the case with constant 1984 prices as base. We choose this year as the base due to the availability of the earliest cross-section in this year. However, this year also happens to be roughly in the middle of the entire 75-year horizon of our data. The figure compares the values of annual growth bias under two approaches: first, when we use the 1984 cross-section for the entire period, and second, when we use this cross-section for all periods before and up to 1984 and use available data on the following cross-sections for all periods between 1984 and 2019. We find it reassuring for our pre-1984 analysis that the bias (and the resulting correction) are similar for both approaches for the post-1984 period.

Recall that the covariance between inflation and income elasticities are negative both before and after 1984 (Figure 13a). As a result, when compared to real consumption exprssed in 1984 prices, the bias in the uncorrected measure of growth changes sign before and after 1984. Figure 14a shows that the annual (cumulative) bias falls from +4.3% (+3.0%) of the uncorrected measured growth in 1954 to around -5.7% (-2.7%) by 2019. In contrast, when we use the prices in 2019 as our base, the annual (cumulative) bias of measured growth is positive over the entire period, reaching to 9.6% (6.9%) by 1954.

Figure 14b shows the corresponding evolution of the corrected *level* of aggregate real consumption relative to the uncorrected measure.[28] When expressed in constant 1984 prices, standard indices of real consumption underestimate aggregate real consumption in 1954 and 2019 by around 2.0% and 0.6%, respectively. In contrast, when we express real consumption in constant 2019 prices, the standard indices overestimate the level of real consumption in 1954 by around 6.3%.

Expressed in terms of real consumption *growth*, from the perspective of prevailing prices in 2019, the standard measures based on reduced-form geometric price indices overestimate (per capita) real consumption growth over the past 75 years by around 14.3 percentage points (or 22 basis points annually). The majority of this overestimation comes from the first half of the data (1954-1984) that contains faster growth rates in the early postwar era. The cumulative overestimation over this 30-year period is around 10 percentage points (or 32 basis points annually).

The corrected and uncorrected measures of growth appear much closer if real consumption is expressed from the perspective of 1984 prices, in which case the bias before and after 1984 changes sign, leading to on overestimation of only 3.1 percentage points over the entire 75-year period (or 5 basis points annually). However, the uncorrected measures contain more substantial error if we use the same constant vector of 1984 prices to compare real consumption between other periods, e.g., between 1954 and 1984. Over this shorter 30-year period, the overestimation of growth is even larger than that over the entire 75-year period, reaching 3.7 percentage points (or 12 basis points annually). In other words, the uncorrected measures do not allow us to perform consistent quantitative comparisons of welfare between arbitrary periods.

## 4.3 Robustness

Using the two alternative datasets previously described, we show that the previous results are robust to data construction and aggregation choices: the size of the nonhomotheticity correction is small with modest inflation and income growth, in particular over the past three decades in the US, but grows substantially when we consider large inflation and growth regimes, as is the case when we extend the analysis to include the entire postwar US experience. All the figures are included in Appendix C.

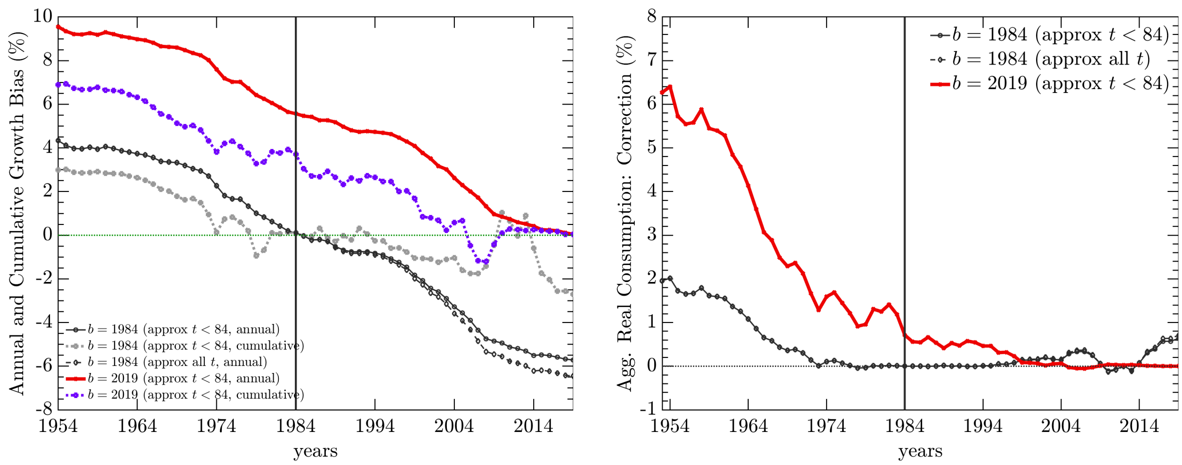
**Official CPI Expenditure Weights** First, we repeat the analysis using the aggregate category-level consumption weights used by the Bureau of Labor Statistics in the official CPI index. Using this dataset, the relationship relationship between household income is negative (Figure A8a), as in our baseline dataset 8a. We find that the higher level of aggregation in this alternative dataset

[28] Figures A7a and A7b in Appendix C presents the same graph using a quadratic approximation for fitting the cross-sectional variations in price indices

Figure 14: CEX and BLS Data: Correction Over Longer Horizon

(a) Nonhomotheticity Growth Bias

(b) Real Consumption Correction




*Note:* Panels (a) and (b) show the evolution of the nonhomotheticity correction and the corrected relative to uncorrected index of real consumption using aggregate consumption expenditure shares.

attenuates both measured real consumption growth (FigureA8b ) and the cross-sectional variations in the reduced-form price indices across household groups, but they still remain decreasing in household income. Figures A9a-A10b show the evolution of the nonhomotheticity bias and the corrected measures of real consumption across different quintiles over the period 1984-2019 with the robustness dataset, which are somewhat smaller than those we found in Figures 9a-11b with the baseline dataset. When we compute the evolution of average real consumption over the period, we find smaller corrections for both initial (1984) and final (2019) base periods (Figures A11a-A11b). Despite the smaller magnitudes due to the more aggregate nature of the BLS data, the overall patterns are broadly similar between the two datasets.

When we consider the longer horizon of the postwar US experience in the robustness dataset, we again find sizable corrections (Figures A12a and A12a). Depending on whether we express real consumption in constant 1984 or 2019 prices, the conventional measures overestimate annual rate of growth in real consumption by 11.3 and 12.4 basis points, respectively.

**Nielsen Scanner Data** Second, we consider the Nielsen data as another robustness check to examine the nonhomotheticity correction in a setting with more disaggregated consumption data. The data covers a much shorter time horizon (2004-2014), but the annual level of inflation inequality is stronger, with substantially lower levels of inflation for higher income households (Figure A13). Even though the error in the conventional reduced-form indices remains fairly small due to minor growth in real consumption, the magnitude of the annual correction reaches 4% of the conventional quantity index after only a decade (Figures A14a-A15b). Overall, these results indicate the robustness of the findings obtained with our baseline dataset.

# 5 Extension: Welfare and Index Numbers with Non-Income Observed Heterogeneity

In this section, we extend our analysis to a setting that includes additional sources of observed heterogeneity among households beyond income. Assume that we observe a vector of characteristics (covariate) $\boldsymbol{\theta}_t^n \in \mathbb{R}_+^D$ for each household $n$ at time $t$.[29] The examples of such characteristics include the age and education of the consumer, the number of household members, and the location of the customer in space. As in Assumption 1, we assume a well-behaved utility or aggregator function that now also depends on the household characteristics $u = U(\boldsymbol{q};\boldsymbol{\theta})$ and let $y = E(u;\boldsymbol{p};\boldsymbol{\theta})$ denote the corresponding expenditure function.

## 5.1 Structural Indices under Observed Heterogeneity

Before constructing structural indices, we need to discuss a conceptual problem in welfare comparisons across households with heterogeneous preferences. Recall that Definition 2 allows us to construct a money metric for the household utility that allows us to compare the welfare of households with homogenous preferences. This money metric for utility corresponds to the unique cardinalization of utility in Equation (6), which accounts for the expenditure required for achieving that level of utility under constant prices. This definition resolves the inherent indeterminacy that all monotonic transformations of utility lead to the same underlying preferences. As we already discussed, these money metrics also correspond to standard measures such as compensating and equivalent variations depending on the choices of base prices.

Consider now a household that moves from an initial vector of quantities and characteristics $(\boldsymbol{q}_{t_0}, \boldsymbol{\theta}_{t_0})$ to the final ones $(\boldsymbol{q}_t, \boldsymbol{\theta}_t)$. First, we want to define a suitable generalization of the notion of cost-of-living to this setting. The most natural approach would be to compare the expenditure required to reach a baseline level of utility under a baseline vector of characteristics $\boldsymbol{\theta}_b$. If we set this level of utility and characteristics to the initial setting for the household, we may again interpret the resulting CoL index as the compensating variations that corresponds to the adjustments in expenditure needed for the household to achieve the original level of utility, *assuming that they still evaluate outcomes based on their original preferences*. If we set them to the final setting, we may interpret the resulting CoL index as the equivalent variations, *assuming that they evaluate outcomes based on their final preferences*.

[29] The assumption that the elements of the vector are positive valued is without loss of generality, as we can always transform the characteristic space in such a way that this condition holds.

To generalize the ReC index, the key question is how to compare welfare under two distinct preferences corresponding to the two vectors of characteristics $\boldsymbol{\theta}_{t_0}$ and $\boldsymbol{\theta}_t$. Under a base vector of prices $\boldsymbol{p}_b$, there is a natural way to compare across such distinct levels of utility $u_{t_0}$ and $u_t$ corresponding to the two vector of characteristics $\boldsymbol{\theta}_{t_0}$ and $\boldsymbol{\theta}_t$, respectively. We can say $u_t$ gives household with preferences $\boldsymbol{\theta}_t$ a higher level of welfare compared to utility $u_{t_0}$ for household with preferences $\boldsymbol{\theta}_{t_0}$ if and only if

$$E\left(u_{t_0};\boldsymbol{p}_b,\boldsymbol{\theta}_{t_0}\right) < E\left(u_t;\boldsymbol{p}_b,\boldsymbol{\theta}_t\right),$$

that is, when the former household needs a higher level of expenditure compared to the latter under the base vector of prices $\boldsymbol{p}_b$. This definition naturally leads us to define real consumption as a unique monotonic transformation (cardinalization) of the utility function

$$q = G_b\left(u;\boldsymbol{\theta}\right) \equiv E\left(u;\boldsymbol{p}_b,\boldsymbol{\theta}\right), \tag{43}$$

that varies based on the characteristics of the household. Therefore, we can say the real income of the household with preferences $\boldsymbol{\theta}_t$ with utility $u_t$ is higher than that of household with preferences $\boldsymbol{\theta}_{t_0}$ and utility $u_{t_0}$ by $q_t - q_{t_0} \equiv G_b\left(u_t;\boldsymbol{\theta}_t\right) - G_b\left(u_{t_0};\boldsymbol{\theta}_{t_0}\right)$ under the base vector of prices $\boldsymbol{p}_b$.

As before, we define an expenditure function in terms of the real consumption (cardinalized utility) as

$$\widetilde{E}_b\left(q;\boldsymbol{p},\boldsymbol{\theta}\right) \equiv E\left(G_b^{-1}\left(q;\boldsymbol{\theta}\right);\boldsymbol{p},\boldsymbol{\theta}\right), \tag{44}$$

where we now have that $\widetilde{E}_b\left(q;\boldsymbol{p}_b,\boldsymbol{\theta}\right) \equiv q$ for all $\boldsymbol{\theta}$. As before, Shephard's lemma implies

$$s_{it} = \Omega_i\left(q_t;\boldsymbol{p}_t,\boldsymbol{\theta}_t\right) \equiv \frac{\partial \log \widetilde{E}_b\left(q_t;\boldsymbol{p}_t,\boldsymbol{\theta}_t\right)}{\partial \log p_{it}}, \tag{45}$$

where the function $\Omega_i(q;\boldsymbol{p}_t,\boldsymbol{\theta}_t)$ is the Hicksian demand function specifying the expenditure share for product $i$ under prices $\boldsymbol{p}_t$ for a consumer with real consumption $q$ and characteristics $\boldsymbol{\theta}_t$ at time $t$.

Given the discussion above, we can now introduce our two structural indices under observable preference heterogeneity: the index of the cost-of-living and the index of real consumption.

**Definition 3.** *Structural Price and Quantity Indices with Observable Heterogeneity.* Consider preferences characterized by utility and expenditure functions $U(\cdot;\boldsymbol{\theta})$ and $E(\cdot;\boldsymbol{\theta})$, respectively. A cost-of-living (CoL) price index for base vectors of quantities $\boldsymbol{q}_b$ and characteristics $\boldsymbol{\theta}_b$ is defined as[30]

$$\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b,\boldsymbol{\theta}_b\right)\equiv\frac{E\left(U\left(\boldsymbol{q}_b;\boldsymbol{\theta}_b\right);\boldsymbol{p}_t,\boldsymbol{\theta}_b\right)}{E\left(U\left(\boldsymbol{q}_b;\boldsymbol{\theta}_b\right);\boldsymbol{p}_{t_0},\boldsymbol{\theta}_b\right)}. \tag{46}$$

A real consumption (ReC) quantity index for a base vector of prices $\boldsymbol{p}_b$ is defined as

$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t,\boldsymbol{\theta}_{t_0},\boldsymbol{\theta}_t;\boldsymbol{p}_b\right)\equiv\frac{E\left(U\left(\boldsymbol{q}_t;\boldsymbol{\theta}_t\right);\boldsymbol{p}_b,\boldsymbol{\theta}_t\right)}{E\left(U\left(\boldsymbol{q}_{t_0};\boldsymbol{\theta}_{t_0}\right);\boldsymbol{p}_b,\boldsymbol{\theta}_{t_0}\right)}. \tag{47}$$

The CoL index characterizes the relative change in the expenditure needed to achieve a certain level of utility for a certain vector of characteristics, as consumers move from the initial vector of prices $\boldsymbol{p}_0$ to the final one $\boldsymbol{p}_1$. The corresponding level of utility is determined using a base vector of quantities $\boldsymbol{q}_b$ and a base vector of characteristics $\boldsymbol{\theta}_b$. This definition fixes the preferences and expresses the changes in the cost-of-living from the perspective of those preferences. In contrast, the ReC index compares the relative change in utility achieved as the vector of characteristics also changes in the two periods, in terms of the expenditure under a fixed vector of base prices $\boldsymbol{p}_b$. Note that by definition, the ReC index gives us the growth in real consumption, that is, $\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t,\boldsymbol{\theta}_{t_0},\boldsymbol{\theta}_t;\boldsymbol{p}_b\right)\equiv G_b\left(U\left(\boldsymbol{q}_t;\boldsymbol{\theta}_t\right);\boldsymbol{\theta}_t\right)/G_b\left(U\left(\boldsymbol{q}_{t_0};\boldsymbol{\theta}_{t_0}\right);\boldsymbol{\theta}_{t_0}\right)$.

Let us investigate our two structural indices under two special case. First, if household preferences do not change, i.e., $\boldsymbol{\theta}_t\equiv\boldsymbol{\theta}_{t_0}$, then both definitions above are reduced to the previous definition 2 that we introduced under homogenuous preferences. Second, if the prices do not change, i.e., $\boldsymbol{p}_t\equiv\boldsymbol{p}_{t_0}$, the CoL price index is always unity, $\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_{t_0};\boldsymbol{q}_b,\boldsymbol{\theta}_b\right)\equiv 1$, and both ReC quantity indices in terms of initial and final periods simply account for the growth in nominal income, $\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_{t_0},\boldsymbol{\theta}_0,\boldsymbol{\theta}_1;\boldsymbol{p}_{t_0}\right)\equiv y_1/y_0$.

## 5.2 Structural and Reduced Indices under Observed Heterogeneity

Given the definitions above, we are now ready to present a number of results that approximate the changes in the ReC index for households in terms of a constant vector of base prices $\boldsymbol{p}_b$. We need an additional assumption on the changes in the characteristics of households.

[30] Again, we can also define the CoL quantity index $\mathscr{Q}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{q}_b,\boldsymbol{\theta}_b\right)$ and ReC price index $\mathscr{P}_{CL}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t,\boldsymbol{\theta}_{t_0},\boldsymbol{\theta}_t;\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{p}_b\right)$ using the corresponding weak factor reversal condition (1). Note that the Laspeyres and Paasche ReC indices coincide with Paasche and Laspeyres CoL quantity indices, respectively:

$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t,\boldsymbol{\theta}_{t_0},\boldsymbol{\theta}_t;\boldsymbol{p}_{t_0}\right)\equiv\mathscr{Q}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{q}_t,\boldsymbol{\theta}_t\right),$$
$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t,\boldsymbol{\theta}_{t_0},\boldsymbol{\theta}_t;\boldsymbol{p}_t\right)\equiv\mathscr{Q}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{q}_{t_0},\boldsymbol{\theta}_{t_0}\right).$$

**Assumption 4.** *Assume that for $t \geq 0$, the maximum change in the logarithm of the characteristics across households in bounded above by a constant $\Delta_\theta \ll 1$ such that*

$$\max_{1\leq d\leq D,\, n\in\mathcal{N}} \left| \log\left(\frac{\theta^n_{d,t+1}}{\theta^n_{d,t}}\right)\right| \leq \Delta_\theta.$$

As before, we again introduce a simplified notation for the CoL price index for a base level of real consumption $q$ under preference characteristics $\boldsymbol{\theta}$, which corresponds to the CoL index for a base basket of quantities that leading to expenditure $q$ under base prices $\boldsymbol{p}_b$ and preference $\boldsymbol{\theta}$:

$$\widetilde{\mathscr{P}}^b_{CL}(\boldsymbol{p}_b, \boldsymbol{p}_t; q, \boldsymbol{\theta}) \equiv \mathscr{P}_{CL}\left(\boldsymbol{p}_b, \boldsymbol{p}_t; \boldsymbol{q}^*_b(q;\boldsymbol{\theta}), \boldsymbol{\theta}\right), \tag{48}$$

where the vector of quantities $\boldsymbol{q}^*_b(q;\boldsymbol{\theta})$ satisfies $q = E\left(U\left(\boldsymbol{q}^*_b(q;\boldsymbol{\theta}); \boldsymbol{\theta}\right); \boldsymbol{p}_b; \boldsymbol{\theta}\right)$.

Now, we are ready to state Lemma 5 below, which generalizes Lemma 2 to the case of heterogeneous preferences, providing a first-order approximation to the ReC index defined by Equation (47).

**Lemma 5.** *If Assumption 4 holds at some period $t$ and if the expenditure function $\log \widetilde{E}_b(\cdot;\cdot,\cdot)$ is continuously differentiable of order at least 2 in all its arguments, the ReC index under the base vector of prices $\mathbf{p}_b$ for each household is related to the CoL price index of that household to the first order of approximation through*

$$\log \mathscr{Q}_{RC}\left(\boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}, \boldsymbol{\theta}^n_t, \boldsymbol{\theta}^n_{t+1}; \boldsymbol{p}_b\right) = \frac{1}{1+\Lambda_b\left(q^n_t; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}^n_{t+1}\right)}\left[\log\left(\frac{y^n_{t+1}/y^n_t}{\mathscr{P}_{CL}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}^n_t, \boldsymbol{\theta}^n_t\right)}\right)\right. \tag{49}$$

$$\left. - \log \mathbb{C}^{(1)}\left(q^n_t; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}^n_t, \boldsymbol{\theta}^n_{t+1}\right)\right] + O\left(\Delta^2\right), \tag{50}$$

*where $\Delta \equiv \max\left\{\Delta_p + \Delta_y, \Delta_\theta\right\}$ and where the nonhomotheticity correction function $\Lambda_b$ is given by*

$$\Lambda_b\left(q; \boldsymbol{p}_t, \boldsymbol{\theta}^n_t\right) \equiv \frac{\partial \log \widetilde{\mathscr{P}}^b_{CL}(\boldsymbol{p}_b, \boldsymbol{p}_t; q, \boldsymbol{\theta}^n_t)}{\partial \log q}, \tag{51}$$

*where the CoL price index is defined in Equation (48). The first-order covariate index $\mathbb{C}^{(1)}$ is given by*

$$\log \mathbb{C}^{(1)}\left(q; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}^n_t, \boldsymbol{\theta}^n_{t+1}\right) \equiv \sum_{d=1}^{D} \Gamma_{b,d}\left(q; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}^n_t\right) \cdot \log\left(\frac{\theta^n_{d,t+1}}{\theta^n_{d,t}}\right), \tag{52}$$

*where the covariate-d correction function $\Gamma_{b,d}$ satisfies:*

$$\Gamma_{b,d}\left(q;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)\equiv\left.\frac{\partial\log\widetilde{\mathscr{P}}_{CL}^{b}\left(\boldsymbol{p}_b,\boldsymbol{p}_t;\boldsymbol{q}_b^*\left(q;\boldsymbol{\theta}\right),\boldsymbol{\theta}\right)}{\partial\log\theta_d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^n}. \tag{53}$$

*Proof.* See Appendix A.3. □

Lemma 5 introduces a covariate index to account for how changes in the preferences of the consumer affect their real consumption. This index aggregates the relative changes in the values of covariates using the values of the corresponding correction functions as weights. Just like the case of the nonhomotheticity correction function in Lemma 3, these covariate correction functions in turn account for the cumulative cross-product covariance between price inflations and the elasticities of demand with respect to each covariate. Lemma 6 below formally introduces this result.

**Lemma 6.** *The covariate-d correction function defined in Equation (53) satisfies:*

$$\Gamma_{b,d}\left(q;\boldsymbol{p}_t,\boldsymbol{\theta}\right)=\int_b^t\left[\sum_{i=1}^{I}\Omega_i\left(q;\boldsymbol{p}_\tau,\boldsymbol{\theta}\right)\zeta_{i,d}\left(q;\boldsymbol{p}_\tau,\boldsymbol{\theta}\right)\frac{d\log p_{i,\tau}}{d\tau}\right]d\tau, \tag{54}$$

*where we have defined the covariate-d elasticity of demand (in terms of expenditure shares) as:*

$$\zeta_{i,d}\left(q;\boldsymbol{p}_t,\boldsymbol{\theta}\right)\equiv\frac{\partial\log\Omega_i\left(q;\boldsymbol{p}_t,\boldsymbol{\theta}\right)}{\partial\log\theta_d}. \tag{55}$$

*Proof.* See Appendix A.3. □

Lemmas 5 and 6 together provide the key insight behind the results in this section. When a given characteristic of households rises, its impact on their real consumption depends on the correlation between price inflation and the elasticity of demand with respect to that characteristic. If this correlation is positive, it implies that, due to the changes in their preferences, the households have to shift their expenditure toward exactly those goods that face higher inflation. For instance, consider the case that households are aging and that price inflation is higher exactly in those products that are highly elastic with respect to the age of the household. The correction in Lemma 5 tells us that, relative to the CoL index of the household under their original age, we need to make additional downward adjustments to our inferred change in the real consumption of households.

To find a first-order approximation of the CoL index $\mathscr{P}_{CL}\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{\theta}_t^n\right)$ in Equation (46), we can simply invoke Equation (12) from Lemma 1 to use the reduced-form geometric price index.[31] Lemma 7 below further provides a second-order approximation that instead relies on the reduced-form Törqvist price index, paralleling the case of Lemma 4 that we saw in the previous section.

**Lemma 7.** *If Assumptions 2 and 4 hold at some period $t$ and if the expenditure function $\log \widetilde{E}_b(\cdot;\cdot,\cdot)$ is continuously differentiable of order at least 3 in all its arguments, the ReC index under the base vector of prices $\boldsymbol{p}_b$ for each household is related to the reduced-form Törqvist price index of that household to the second order of approximation in $\Delta \equiv \max\left\{\Delta_p + \Delta_y, \Delta_\theta\right\}$ through*

$$\log \mathcal{Q}_{RC}\left(\boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n; \boldsymbol{p}_b\right) = \frac{1}{1+\frac{1}{2}\left[\Lambda_b\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{\theta}_t^n\right) + \Lambda_b\left(q_{t+1}^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^n\right)\right]} \tag{56}$$

$$\times\left[\log\left(\frac{\left(y_{t+1}^n / y_t^n\right)}{\mathbb{P}_T\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right)}\right) - \log \mathbb{C}^{(2)}\left(q_t^n, q_{t+1}^n; \boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n\right)\right] + O\left(\Delta^3\right), \tag{57}$$

*where the nonhomotheticity correction function $\Lambda_b\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{\theta}_t^n\right)$ is defined by Equation (20), and where the second-order covariate index $\mathbb{C}^{(2)}$ is given by*

$$\mathbb{C}^{(2)}\left(q_t^n, q_{t+1}^n; \boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n\right) \equiv \frac{1}{2}\sum_{d=1}^{D}\left[\Gamma_{b,d}\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{\theta}_t^n\right) + \Gamma_{b,d}\left(q_{t+1}^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^n\right)\right] \cdot \log\left(\frac{\theta_{d,t+1}}{\theta_{d,t}}\right),$$

*with the covariate-$d$ correction function $\Gamma_{b,d}$ satisfying Equation (53).*

*Proof.* See Appendix A.3. □

Here, we provide an algorithm that allows us to approximate changes in ReC indices of households to the first order of approximation. The idea of this algorithm closely parallels that of Algorithm 1 in the previous section: we use the cross-sectional variations in the reduced-form price indices to nonparametrically approximate the correction functions for nonhomotheticity and all covariates.

**Algorithm 3.** *Let $\widehat{q}_b^n \equiv q_b^n \equiv y_b^n$ and define a function $\widehat{\mathscr{P}}_b^{(1)}(q, \boldsymbol{\theta}) \equiv 1$, and consider a sequence $\{g_k(q, \boldsymbol{\theta})\}_{k=0}^{K_N}$ of power functions of $\log q$ and $\log \boldsymbol{\theta}$ where $K_N$ depends on $N$, the number of households in the cross-section. For each $t \geq b$, apply the following steps:*

[31] See Lemma 11 in Appendix A.3.

1. *Compute the next period function* $\widehat{\mathscr{P}}^{(1)}_{t+1}(\cdot)$*:*

$$\log \widehat{\mathscr{P}}^{(1)}_{t+1}(q,\boldsymbol{\theta}) \equiv \log \widehat{\mathscr{P}}^{(1)}_{t}(q) + \sum_{k=0}^{K} \widehat{\alpha}_{k,t}\, g_k(q,\boldsymbol{\theta}), \tag{58}$$

*where the coefficients* $\left(\widehat{\alpha}_{k,t}\right)_{k=0}^{K_N}$ *solve the following problem:*

$$\min_{\left(\alpha_{k,t}\right)_{k=0}^{K}} \sum_{n=1}^{N} \left( \log \mathbb{P}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right) - \sum_{k=0}^{K_N} \alpha_{k,t}\, g_k\left(\widehat{q}_t^n, \boldsymbol{\theta}_t^n\right) \right)^2. \tag{59}$$

2. *Compute the real consumption in the next period for each household:*

$$\log \widehat{q}_{t+1}^n = \log \widehat{q}_t^n + \frac{1}{1+\widehat{\Lambda}^{(1)}_{t+1}\left(\widehat{q}_t^n, \boldsymbol{\theta}_t^n\right)} \left[ \log\left( \frac{\left(y_{t+1}^n / y_t^n\right)/}{\mathbb{P}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right)} \right) \right. \tag{60}$$

$$\left. - \log \widehat{\mathbb{C}}^{(1)}\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n\right) \right], \tag{61}$$

*where we have defined the approximate nonhomotheticity correction function as:*

$$\widehat{\Lambda}^{(1)}_{t+1}(q,\boldsymbol{\theta}) \equiv \frac{\partial}{\partial \log q} \log \widehat{\mathscr{P}}^{(1)}_{t+1}(q,\boldsymbol{\theta}) = \sum_{k=0}^{K_N} \left( \sum_{\tau=b+1}^{t+1} \widehat{\alpha}_{k,\tau} \right) \frac{\partial g_k(q,\boldsymbol{\theta})}{\partial \log q}, \tag{62}$$

*and the approximate first-order covariate index is given by*

$$\widehat{\mathbb{C}}^{(1)}\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n\right) \equiv \sum_{d=1}^{D} \widehat{\Gamma}^{(1)}_{d,t+1}\left(q_t^n, \boldsymbol{\theta}_t^n\right) \cdot \log\left( \frac{\theta_{d,t+1}}{\theta_{d,t}} \right)$$

*with the following approximation for the covariate-$d$ correction function:*

$$\widehat{\Gamma}^{(1)}_{d,t+1}(q,\boldsymbol{\theta}) \equiv \frac{\partial}{\partial \log \theta_d} \log \widehat{\mathscr{P}}^{(1)}_{t+1}(q,\boldsymbol{\theta}) = \sum_{k=0}^{K_N} \left( \sum_{\tau=b+1}^{t+1} \widehat{\alpha}_{k,\tau} \right) \frac{\partial g_k(q,\boldsymbol{\theta})}{\partial \log \theta_d}.$$

Proposition 3 establishes bounds on the approximation error of the sequences of ReC indices found by Algorithm 3. The main additional requirement, compared to Proposition 1, is that we now require the expenditure function to be infinitely differentiable.

**Proposition 3.** *Assume that the expenditure function* $\log \widetilde{E}_b(\cdot;\cdot)$ *is an analytical function. If Assumptions 2, 4, and 3 hold, then the sequences of real consumptions* $\boldsymbol{q}^n_{(b,T)}$ *constructed by Algorithm*

3 *satisfy:*

$$\log \mathcal{Q}_{RC}\left(\boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n; \boldsymbol{p}_b\right) = \log\left(\frac{\widehat{q}_{t+1}^n}{\widehat{q}_t^n}\right) + O\left(\Delta^2\right) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}} \cdot \Delta^4 + K_N^{-m}\right)\Delta\right), \tag{63}$$

*where* $\Delta \equiv \max\left\{\Delta_p + \Delta_y, \Delta_\theta\right\}$ *and m is any positive number.*

*Proof.* See Appendix A.3. □

With the stronger assumption imposed by Proposition 3 on the differentiability of the expenditure function, we find a tighter bound in Equation (63). Again, so long as $K_N \to \infty$ and $K_N^7/N \to 0$, the error in our approximation converges to zero. In Appendix A.3, we further provide a generalization of the second-order approximation for the case of observed heterogeneity. Algorithm 4 and Proposition 4 provide generalizations of Algorithm 2 and Proposition 2, respectively, to the cases involving observed heterogeneity.

# 6 Conclusion

In this paper, we extended insights from classical demand theory to the case of observable preference heterogeneity, in particular nonhomothetic preferences. We obtained a procedure for nonparametric measurement of consumer welfare, providing a theoretically consistent measure of real consumption while imposing minimal restrictions on the underlying preferences. This approach remains valid under any observable heterogeneity and requires only data on spending patterns in a cross-section of consumers. We showed the practical relevance of the correction for nonhomotheticities when computing long-run growth in consumer welfare. With our correction taking 2019 prices as base, growth in consumer welfare is significantly attenuated in the United States in the post-war era, due to the combination of fast growth and lower inflation for income-elastic product categories. Extending this analysis to other countries and time periods, as well as to the measurement of purchasing power parity (PPP) indices across countries with preference heterogeneity, is a promising direction for future research.

# References


**Argente, David and Munseob Lee**, "Cost of Living Inequality during the Great Recession," *Journal of the European Economic Association*, 2021, *19* (2), 913–952. 2, 7, 35

**Atkin, David, Benjamin Faber, Thibault Fally, and Marco Gonzale-Navarro**, "Measuring Welfare and Inequality with Incomplete Price Information," 2020. 6

**Baqaee, David and Ariel Burstein**, "Welfare and Output with Income Effects and Demand Instability," 2021. 2, 6, A3

**Battistin, Erich, Richard Blundell, and Arthur Lewbel**, "Why Is Consumption More Log Normal than Income? Gibrat's Law Revisited," *Journal of Political Economy*, 2009, *117* (6), 1140–1154. 25

**Comin, Diego, Danial Lashkari, and Marti Mestieri**, "Structural change with long-run income and price effects," *Econometrica*, 2021, *89* (1), 311–374. 4, 24, 25

**Deaton, Angus and John Muellbauer**, "An almost ideal demand system," *American Economic Revieweview*, 1980, *70* (3), 312–326. 6

**Diewert, W Erwin**, "Exact and superlative index numbers," *Journal of Econometrics*, 1976, *4* (2), 115–145. 3, 6, 9, 13, A2, A3

_ , "Superlative Index Numbers and Consistency in Aggregation," *Econometrica*, 1978, *46* (4), 883–900. 8, 11, A2

**Diewert, Walter E**, "The economic theory of index numbers: a survey," in Walter E Diewert and Alice O Nakamura, eds., *Essays in Index Number Theory*, Emerald Group Publishing Ltd, 1993, chapter 7, pp. 163–208. 6, 8, 9, A1, A2

**Engel, Ernst**, "Die Produktions- und Consumtionsverhältnisse des Königreichs Sachsen," *Zeitschrift des Statistischen Büreaus des Königlich Sächischen Ministeriums des Innern*, 1857, *8-9*, 1–54. 2

**Fajgelbaum, Pablo D and Amit K Khandelwal**, "Measuring the Unequal Gains from Trade," *Quarterly Journal of Economics*, 2016, *131* (3), 1113–1180. 6

**Hobijn, Bart and David Lagakos**, "Inflation inequality in the United States," *Review of Income and Wealth*, 2005, *51* (4), 581–606. 6

**Jaravel, Xavier**, "The unequal gains from product innovations: evidence from the US retail sector," *Quarterly Journal of Economics*, 2019, *134* (2), 715–783. 2, 7, 34, 35

**Kaplan, Greg and Sam Schulhofer-Wohl**, "Inflation at the household level," *Journal of Monetary Economics*, 2017, *91*, 19–38. 2, 7, 35

**McGranahan, Leslie and Anna Paulson**, "Federal Reserve Bank of Chicago Demographic Group : 1983-2005 Inflation Experiences by Demographic Group : 1983-2005 Federal Reserve Bank of Chicago," 2006. 7

**Newey, Whitney K.**, "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 1997, *79* (1), 147–168. 19, A9, A10

**Pollak, Robert A**, "The theory of the cost-of-living index," in W Erwin Diewert, ed., *Price Level Measurement*, Elsevier North Holland Inc., 1990, pp. 5–77. 6, 9

**Redding, Stephen J. and David E. Weinstein**, "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences," *Quarterly Journal of Economics*, 2020, *135* (1), 503–560. 6

**Samuelson, Paul A and S Swamy**, "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *American Economic Review*, 1974, *64* (4), 566–593. 2, 6, 9, A1, A2, A3

# A Theory Appendix

## A.1 Additional Definitions and Results

### A.1.1 Conjugate Structural Indices

First, when comparing welfare only across two periods, the CoL price index and the ReC quantity index are conjugates when the base vectors of quantities and prices belong to opposite periods:

$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_b\right)\times\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{-b}\right)=\frac{y_t}{y_{t_0}},\qquad (b,-b)\in\{(t_0,t),(t,t_0)\}. \tag{A1}$$

For comparisons across multiple periods, we can define a conjugate ReC price index that satisfies weak factor reversal with the ReC quantity index with base period $b$, we find that it is given by[32]

$$\mathscr{P}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{p}_b\right)\equiv\frac{y_t/y_{t_0}}{\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_b\right)}=\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_b;\boldsymbol{q}_{t_0}\right)\times\mathscr{P}_{CL}\left(\boldsymbol{p}_b,\boldsymbol{p}_t;\boldsymbol{q}_t\right). \tag{A2}$$

The expression on the right hand side of the second equality is a composite CoL price index, through which we first evaluate the cost-of-living change from the initial period to the base period, from the perspective of the initial consumption basket, and then evaluate the cost-of-living change from the base period to the final period, now from the perspective of the final consumption basket. Similarly, we can define a CoL quantity index using the weak factor reversal condition:

$$\mathscr{Q}_{CL}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{p}_b\right)\equiv\frac{y_t/y_{t_0}}{\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right)}=\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_b;\boldsymbol{p}_{t_0}\right)\times\mathscr{Q}_{CL}\left(\boldsymbol{q}_b,\boldsymbol{q}_t;\boldsymbol{p}_t\right). \tag{A3}$$

Again, this quantity index corresponds to a composite ReC quantity index in which we first evaluate the real consumption change from the initial period to the base period from the perspective of initial prices, and multiply that with the real consumption change from the base to the final period from the perspective of final period prices. Importantly, note that unlike the case of ReC quantity indices, we cannot necessarily use such a CoL quantity index to consistently compare welfare over time due to the change in the price vector used for the evaluation of welfare.These observations illustrate the importance of holding constant the base vector of prices and quantities to implement principled welfare comparisons over the long run.

[32] Equation (A2) follows from the identity $\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_b\right)\equiv\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_b;\boldsymbol{p}_b\right)\times\mathscr{Q}_{RC}\left(\boldsymbol{q}_b,\boldsymbol{q}_t;\boldsymbol{p}_b\right)$ and Equation (A1).

### A.1.2 Homothetic Preferences

If (and only if) the utility function $U(\cdot)$ is homothetic, then we can write the expenditure function as

$$E(u;\boldsymbol{p}) = P(\boldsymbol{p}) \cdot F(u), \tag{A4}$$

for some unit cost function $P(\cdot)$ and some canonical homothetic cardinalization $F(\cdot)$ of utility (Diewert, 1993).[33] This fact allows us to derive the important results that homotheticity is a necessary and sufficient condition for the values of the ReC and CoL quantity indices to be independent of the base year vectors of prices $\boldsymbol{p}_b$ and quantities $\boldsymbol{q}_b$, respectively, and to also coincide for any pair of periods.[34] To be more precise, if (and only if) the utility function $U(\cdot)$ rationalizing the sequence of prices and quantities $(\boldsymbol{p}_t, \boldsymbol{q}_t)_{t=0}^{T}$ is homothetic, we have:

$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}; \boldsymbol{p}_b\right) = \frac{F\left(u_{t+1}\right)}{F\left(u_t\right)}, \qquad \forall b, \tag{A6}$$

$$\mathscr{P}_{CL}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_b\right) = \frac{P\left(\boldsymbol{p}_{t+1}\right)}{P\left(\boldsymbol{p}_t\right)}, \qquad \forall b, \tag{A7}$$

for all $\boldsymbol{p}_b$ and $\boldsymbol{q}_b$, and where $u_t \equiv U\left(\boldsymbol{q}_t\right)$. Equations (A6) shows that the CoL prices index captures the rise in the unit costs and the ReC quantity index captures the rise in the canonical cardinalization $F(u)$, which can only be defined under the assumption of homotheticity. Moreover, the latter also coincides with the growth in the real consumption cardinalization $G_b(\cdot)$ defined based on Equation (6).

Most importantly, Equations (A6) and (A7) imply that, under homothetic preferences, the structural indices yield the same values independently of the base quantities and prices. Recall from our discussion of Equation (A1) that the measures of growth in real consumption systematically vary in their base vector of prices if the cost-of-living measures vary in consumer income.. Under homotheticity, the inflation in the measures of cost-of-living do not depend on the income of the consumer (as captured by the utility associated with base quantities $\boldsymbol{q}_b$) in Equation (A7). Correspondingly, the measures of real consumption do not depend on the vector of prices chosen to express them in Equation (A6), and ReC quantity indices satisfy the following condition with

[33] From Equation (A4), if the preferences are homothetic we have:

$$\widetilde{E}_b(q;\boldsymbol{p}) = \frac{P(\boldsymbol{p})}{P(\boldsymbol{p}_b)} q. \tag{A5}$$

[34] Samuelson and Swamy (1974) refer to this result as the *homogeneity theorem.*

CoL price indices irrespective of base prices and quantities:

$$\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0},\boldsymbol{q}_t;\boldsymbol{p}_b\right)\times\mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{b'}\right)=\frac{y_t}{y_{t_0}},\qquad \forall b,b'. \tag{A8}$$

## A.2 Connection with Classical Index Number Theory

The standard treatment of inex numbers (Diewert, 1993) defines a quantity index $\mathbb{Q}$ to be *exact* for given family of homothetic preferences characterized by parametric utility functions $U(\cdot;\boldsymbol{\theta})$ if for any sequence of consumption choices $\left(\boldsymbol{p}_{(0,T)};\boldsymbol{q}_{(0,T)}\right)$ rationalized by a member of this family, the value of the quantity index $\mathbb{Q}\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t,\boldsymbol{q}_{t+1}\right)$ is exactly equal to the corresponding ReC quantity index $\mathscr{Q}_{RC}\left(\boldsymbol{q}_t,\boldsymbol{q}_{t+1};\boldsymbol{p}_b\right)$ for any vector of base quantities $\boldsymbol{q}_b$. Similarly, we can define a price index $\mathbb{P}$ to be exact for the same family if it precisely coincides with the value of the corresponding CoL price index for $\mathscr{P}_{CL}\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_b\right)$ for any base basket of consumption $\boldsymbol{q}_b$. For instance, the Törqvist price index is exact for preferences that lead to a translog unit cost function.[35] If the price index $\mathbb{P}$ is exact for a family of preferences, then the chained price index over multiple periods also yields the value of the CoL or ReC price index over longer horizons, e.g., $\mathscr{P}_{CL}(\boldsymbol{p}_0,\boldsymbol{p}_T;\boldsymbol{q}_b)=\mathbb{P}^c\left(\boldsymbol{p}_{(0,T)};\boldsymbol{q}_{(0,T)}\right)$ for any vector of base quantities $\boldsymbol{q}_b$, where the chained index is defined through Equation (10).

Needless to say, the concept of exact price indices still requires us to take a stance on the underlying form of the preference functions. One crucial step is to define, as in Diewert (1976), the ideal and Törqvist price indices as *superlative* price indices, on the grounds that they are exact for families of preferences, e.g., the translog family, that can provide a second-order approximation to other homothetic preferences. In line with this insight, Diewert (1978) has shown that alternative choices of superlative indices, when chained, lead to very similar estimates for the changes in cost-of-living and real consumption in practice. Lemma 1 formalizes these classical insights from a different angle. Instead of establishing the exactness of the reduced-form indices for a family of preferences that may approximate general preferences, our Lemma 1 above instead provides bounds on the approximation error of the reduced-form indices for general preferences.

Many important results of the index number theory do not generalize beyond the case of homothetic preferences.[36] Crucially, the equivalence between the ReC and CoL indices high-

[35] As for other examples, the Laspeyres and Paasche indices are exact for Leontief utility functions, while the geometric index is exact for Cobb-Douglas utility functions. The ideal price index is exact for the family of preferences that lead to quadratic unit cost functions.

[36] Samuelson and Swamy (1974) discuss several examples of such results and provide examples that show how they fail under nonhomotheticity. Examples include the result that Laspeyres and Paasche indices provide bounds for the ReC and CoL indices:

$$\mathbb{P}_L\le\mathscr{P}_{CL}\le\mathbb{P}_P,\qquad \mathbb{Q}_P\le\mathscr{Q}_{RC}\le\mathbb{Q}_L,$$

as well as the result that the Divisia indices provide a unique decomposition of the changes in total expenditure.

lighted in Equation (A6) breaks down under nonhomotheticity. The most important extensions of the index number theory to the case of nonhomothetic preferences involve CoL indices. For instance, Diewert (1976) shows that the reduced-form Törqvist price index is exact for the family of nonhomothetic preferences characterized by a translog expenditure function in the following restricted sense: the reduced-form index in this case is exact only for a specific CoL index $\mathscr{P}_{CL}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}^*\right)$ with the base basket of consumption $\boldsymbol{q}_t^*$ that satisfies $U(\boldsymbol{q}^*) \equiv \sqrt{U\left(\boldsymbol{q}_{t_0}\right) \times U(\boldsymbol{q}_t)}$. In other words, the superlative price index provides us with the change in the cost-of-living for a consumer that aims to achieve a level of utility in between that of the initial and the final periods. Our Lemma 1 above offers an alternative rendition of this insight by providing bounds on the approximation error of the reduced-form index for a basket of quantities $\boldsymbol{q}_b^*$ such that the corresponding real consumption, as defined by Equation (6), satisfies $G_b\left(U\left(\boldsymbol{q}_b^*\right)\right) \equiv \sqrt{G_b\left(U\left(\boldsymbol{q}_{t_0}\right)\right) \times G_b\left(U(\boldsymbol{q}_t)\right)}$.

As discussed in Se_bction 2.2, while the results of Diewert (1976) appear reassuring, they do not provide any justification for relying on the chained reduced-form Törqvist price index, defined by Equation (10), to compute the CoL price index over long horizons. The reason is simply that the base basket of quantities $\boldsymbol{q}_t^*$ corresponding to the CoL index changes along the chain. As Samuelson and Swamy (1974) and more recently Baqaee and Burstein (2021) have shown, we can still compute the structural indices over longer horizons if we have access to the estimates of elasticities of substitution in preferences. However, in order to estimate the elasticities of substitution we need to first specify a particular class of preferences. In the next section, we lay out a number of results that allow us to extend the classical index number theory to the case of real consumption indices without the need to such estimation.

## A.3 Proofs and Derivations

**Lemma.** *See Lemma 1.*

*Proof.* Performing of a Taylor series expansion of the expenditure function around the vector of prices $\boldsymbol{p}_{t_0}$, we find:

$$\begin{aligned}\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_{t_0}\right) &= \log \frac{E\left(u_{t_0}; \boldsymbol{p}_t\right)}{E\left(u_{t_0}; \boldsymbol{p}_{t_0}\right)}, \\ &= \sum_{i=1}^{I} s_{i,t_0} \log\left(\frac{p_{i,t}}{p_{i,t_0}}\right) + O\left(|t-t_0|^2 \Delta_p^2\right),\end{aligned} \tag{A9}$$

where we have used Shephard's lemma in the second equality, leading to Equation (12) from definition of the geometric price index (2).

Using Lemma 8 below, we can write the Laspeyres CoL index from Equation (4) as:

$$\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_{t_0}\right) = \log \frac{E\left(u_{t_0}; \boldsymbol{p}_t\right)}{E\left(u_{t_0}; \boldsymbol{p}_{t_0}\right)},$$

$$= \frac{1}{2}\sum_{i=1}^{I}\left[\Omega_i\left(u_{t_0}; \boldsymbol{p}_{t_0}\right) + \Omega_i\left(u_{t_0}; \boldsymbol{p}_t\right)\right]\log\left(\frac{p_{i,t}}{p_{i,t_0}}\right) + O\left(|t-t_0|^3 \Delta_p^3\right), \quad \text{(A10)}$$

where $\Omega_i(u; \boldsymbol{p})$ is the share of good-$i$ in total expenditure based on the corresponding Hicksian demand function with utility $u$ at prices $\boldsymbol{p}$. From homotheticity, we know that $\Omega_i\left(u_{t_0}; \boldsymbol{p}_t\right) = \Omega_i(u_t; \boldsymbol{p}_t) = s_{i,t}$, which, combined with definition (3), leads to the desired result.

For the case of nonhomothetic preferences, using the definition of (7), we can write:

$$\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_b\right) = \log \frac{\widetilde{E}_b\left(q_b^b; \boldsymbol{p}_t\right)}{\widetilde{E}_b\left(q_b^b; \boldsymbol{p}_{t_0}\right)},$$

$$= \sum_{i=1}^{I} \Omega_i^b\left(q_b^b; \boldsymbol{p}_{t_0}\right) \log\left(\frac{p_{i,t}}{p_{i,t_0}}\right) + O\left(|t-t_0|^2 \Delta_p^2\right). \quad \text{(A11)}$$

Performing of a Taylor series expansion of the expenditure function around the vector of prices $\boldsymbol{p}_{t_0}$, we find:

$$\Omega_i^b\left(q_b; \boldsymbol{p}_{t_0}\right) = \Omega_i^b\left(q_{t_0}; \boldsymbol{p}_{t_0}\right) + \frac{\partial \Omega_i^b\left(q_{t_0}; \boldsymbol{p}_{t_0}\right)}{\partial \log q} \cdot \log\left(\frac{q_b}{q_{t_0}}\right) + O\left(|b-t_0|^2 \Delta_q^2\right),$$

where $\Delta_q$ stands for a bound on the change in log real consumption. Substituting this expression in Equation (A11), we find:

$$\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_b\right) = \log \mathbb{P}_G\left(\boldsymbol{p}_{t_0}, \boldsymbol{p}_t; \boldsymbol{q}_{t_0}^n, \boldsymbol{q}_t^n\right) + \log\left(\frac{q_b}{q_{t_0}}\right) \cdot \sum_{i=1}^{I} \frac{\partial \Omega_i^b\left(q_{t_0}; \boldsymbol{p}_{t_0}\right)}{\partial \log q} \log\left(\frac{p_{i,t}}{p_{i,t_0}}\right)$$
$$+ O\left(|b-t_0|^2 \Delta_q^2 \cdot |t-t_0| \Delta_p\right) + O\left(|t-t_0|^2 \Delta_p^2\right)$$

Since $\widetilde{E}_b(q_b; \boldsymbol{p}_t)$ is second order continuously differentiable, the second term above is of the order $O\left(|t-t_0|\Delta_p \cdot |b-t_0|\Delta_q\right)$. Invoking Lemma 9 allows us to bound $\Delta_q \leq \Delta_y + \Delta_p$ and leads to Equation (14) since $|b-t_0| \leq |t-t_0|$.

Using Lemma 8 below and the definition of (7), we can write the CoL index from Equation

(4) as:

$$\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right) = \log \frac{\widetilde{E}_b\left(q_b^b;\boldsymbol{p}_t\right)}{\widetilde{E}_b\left(q_b^b;\boldsymbol{p}_{t_0}\right)},$$

$$= \frac{1}{2}\sum_{i=1}^{I}\left[\Omega_i^b\left(q_b^b;\boldsymbol{p}_{t_0}\right)+\Omega_i^b\left(q_b^b;\boldsymbol{p}_t\right)\right]\log\left(\frac{p_{i,t}}{p_{i,t_0}}\right)+O\left(|t-t_0|^3\,\Delta_p^3\right). \quad \text{(A12)}$$

Again, using Lemma 8 on the Hicksian demand function we find:

$$\Omega_i^b\left(q_b;\boldsymbol{p}_{t_0}\right) = \Omega_i^b\left(q_{t_0};\boldsymbol{p}_{t_0}\right)+\frac{1}{2}\left[\frac{\partial \Omega_i^b\left(q_{t_0}^b;\boldsymbol{p}_{t_0}\right)}{\partial \log q}+\frac{\partial \Omega_i^b\left(q_b^b;\boldsymbol{p}_{t_0}\right)}{\partial \log q}\right]\cdot\log\left(\frac{q_b^b}{q_{t_0}^b}\right)+O\left(|b-t_0|^3\cdot\Delta_q^3\right),$$

$$\Omega_i^b\left(q_b;\boldsymbol{p}_t\right) = \Omega_i^b\left(q_t;\boldsymbol{p}_t\right)+\frac{1}{2}\left[\frac{\partial \Omega_i^b\left(q_t^b;\boldsymbol{p}_t\right)}{\partial \log q}+\frac{\partial \Omega_i^b\left(q_b^b;\boldsymbol{p}_t\right)}{\partial \log q}\right]\cdot\log\left(\frac{q_b^b}{q_t^b}\right)+O\left(|b-t_0|^3\cdot\Delta_q^3\right)$$

Substituting this expression in Equation (A11), invoking Lemma 9 below, we find:

$$\begin{aligned}\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right) &= \log \mathbb{P}_T\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0}^n,\boldsymbol{q}_t^n\right)\\ &+\frac{1}{2}\cdot\log\left(\frac{q_b^b}{q_{t_0}^b}\right)\cdot\sum_{i=1}^{I}\left[\frac{\partial \Omega_i^b\left(q_{t_0}^b;\boldsymbol{p}_{t_0}\right)}{\partial \log q}+\frac{\partial \Omega_i^b\left(q_b^b;\boldsymbol{p}_{t_0}\right)}{\partial \log q}\right]\log\left(\frac{p_{i,t}}{p_{i,t_0}}\right)\\ &+\frac{1}{2}\cdot\log\left(\frac{q_b^b}{q_t^b}\right)\cdot\sum_{i=1}^{I}\left[\frac{\partial \Omega_i^b\left(q_t^b;\boldsymbol{p}_t\right)}{\partial \log q}+\frac{\partial \Omega_i^b\left(q_b^b;\boldsymbol{p}_t\right)}{\partial \log q}\right]\log\left(\frac{p_{i,t}}{p_{i,t_0}}\right)+O\left(|b-t_0|^3\,\Delta_q^3\right).\end{aligned} \quad \text{(A13)}$$

Now, we use the third-order continuously differentiable property of the expenditure function to find

$$\frac{\partial \Omega_i^b\left(q_{t'}^b;\boldsymbol{p}_{t''}\right)}{\partial \log q} = \frac{\partial \Omega_i^b\left(q_{t_0}^b;\boldsymbol{p}_{t_0}\right)}{\partial \log q}+O\left(|t-t_0|\,\Delta\right), \qquad t',t''\in[t_0,t], \qquad \Delta\equiv\max\left\{\Delta_p,\Delta_q\right\},$$

and to substitute for the expressions within the square brackets in Equation (A13). Since $|b-t_0|\le|t-t_0|$, this leads to the following result

$$\begin{aligned}\log \mathscr{P}_{CL}\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_b\right) &= \log \mathbb{P}_T\left(\boldsymbol{p}_{t_0},\boldsymbol{p}_t;\boldsymbol{q}_{t_0}^n,\boldsymbol{q}_t^n\right)\\ &+\log\left[\frac{\left(q_b^b\right)^2}{q_{t_0}^b\cdot q_t^b}\right]\cdot\sum_{i=1}^{I}\frac{\partial \Omega_i^b\left(q_{t_0}^b;\boldsymbol{p}_{t_0}\right)}{\partial \log q}\log\left(\frac{p_{i,t}}{p_{i,t_0}}\right)+O\left(|t-t_0|^3\,\Delta^3\right).\end{aligned}$$

Thus, if we let $\frac{\left(q_b^b\right)^*}{q_{t_0}^b} = \frac{q_t^b}{\left(q_b^b\right)^*}$, implying the condition $\mathscr{Q}_{RC}\left(\boldsymbol{q}_{t_0}, \boldsymbol{q}_b^*; \boldsymbol{p}_b\right) = \mathscr{Q}_{RC}\left(\boldsymbol{q}_b^*, \boldsymbol{q}_t; \boldsymbol{p}_b\right)$ in the statement of the lemma, the second term on the right hand side vanishes. Using Lemma 9 below, we reach the desired result. □

**Lemma.** *See Lemma 2.*

*Proof.* First, note that we have:

$$\log\left(\frac{y_t}{y_{t_0}}\right) = \log\frac{\widetilde{E}_b\left(q_t^b; \boldsymbol{p}_t\right)}{\widetilde{E}_b\left(q_{t_0}^b; \boldsymbol{p}_{t_0}\right)}.$$

We can do a first-order Taylor expansion of the left-hand-side of the equation above in terms of $q_t^b$, assuming that $\left|\log\left(\frac{q_t^n}{q_{t_0}^n}\right)\right| < \Delta_q$:

$$\log\left(\frac{y_t}{y_{t_0}}\right) = \log\frac{\widetilde{E}_b\left(q_{t_0}^b; \boldsymbol{p}_t\right)}{\widetilde{E}_b\left(q_{t_0}^b; \boldsymbol{p}_{t_0}\right)} + \left.\frac{\partial \log\widetilde{E}_b\left(q; \boldsymbol{p}_t\right)}{\partial \log q}\right|_{q \equiv q_{t_0}^b} \cdot \log\left(\frac{q_t^b}{q_{t_0}^b}\right) + O\left(\Delta_q^2\right).$$

This immediately leads to Equation (19), using definitions (4) and (5).

Next, note that from Equation A5, we have $\Lambda_b\left(q; \boldsymbol{p}_b\right) \equiv 1$. This allows us to write:

$$\Lambda_b\left(q; \boldsymbol{p}_t\right) = \frac{\partial}{\partial \log q} \log\left(\frac{\widetilde{E}_b\left(q; \boldsymbol{p}_t\right)}{\widetilde{E}_b\left(q; \boldsymbol{p}_b\right)}\right), \tag{A14}$$

which, using the definition of the CoL price index in Equation (4), yields Equations (19) and (20). □

**Lemma 8.** *Consider a function $f(\boldsymbol{x})$ defined in the space of $\boldsymbol{x} \in \mathbb{R}^I$. To the second order of approximation, we have:*

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{I}\left[\frac{\partial f(\boldsymbol{y})}{\partial y_i} + \frac{\partial f(\boldsymbol{x})}{\partial x_i}\right]\left(y_i - x_i\right).$$

*Proof.* Using Taylor's expansion, up to the second order in $\boldsymbol{y} - \boldsymbol{x}$, we have:

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \sum_{i=1}^{I}\frac{\partial f(\boldsymbol{x})}{\partial x_i}\left(y_i - x_i\right) + \frac{1}{2}\sum_{i,j=1}^{I}\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}\left(y_i - x_i\right)\left(y_j - x_j\right),$$

$$f(\boldsymbol{x}) = f(\boldsymbol{y}) + \sum_{i=1}^{I}\frac{\partial f(\boldsymbol{y})}{\partial x_i}\left(x_i - y_i\right) + \frac{1}{2}\sum_{i,j=1}^{I}\frac{\partial^2 f(\boldsymbol{y})}{\partial x_i \partial x_j}\left(y_i - x_i\right)\left(y_j - x_j\right).$$

Together, the two equations imply:

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \frac{1}{2}\sum_{i=1}^{I}\left[\frac{\partial f(\boldsymbol{y})}{\partial y_i} + \frac{\partial f(\boldsymbol{x})}{\partial x_i}\right](y_i - x_i) + \frac{1}{4}\sum_{i,j=1}^{I}\left[\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} - \frac{\partial^2 f(\boldsymbol{y})}{\partial x_i \partial x_j}\right](y_i - x_i)\left(y_j - x_j\right).$$

This gives us the desired result since:

$$\frac{\partial^2 f(\boldsymbol{y})}{\partial x_i \partial x_j} - \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \sum_k \frac{\partial^3 f(\boldsymbol{x})}{\partial x_k \partial x_i \partial x_j}(y_k - x_k).$$

□

**Lemma 9.** *If Assumptions 1 and 2 hold, then we have:*

$$\Delta_q \equiv \max_{0\le t\le T}\left|\log\left(\frac{q_{t+1}^b}{q_t^b}\right)\right| = O\left(\Delta_p + \Delta_y\right), \tag{A15}$$

*where $q_t^b \equiv y_b \times \mathcal{Q}_{RC}(\boldsymbol{q}_b, \boldsymbol{q}_t; \boldsymbol{p}_b)$ stands for real consumption expressed in terms of any base period $b$.*

*Proof.* Using a Taylor expansion of $\widetilde{E}_b(q_{t+1}^b; \boldsymbol{p}_{t+1})$ around $q_t^b$ and $\boldsymbol{p}_t$, we find

$$\begin{aligned}\log\left(\frac{y_{t+1}}{y_t}\right) &= \log\left(\frac{\widetilde{E}_b(q_{t+1}^b; \boldsymbol{p}_{t+1})}{\widetilde{E}_b(q_t^b; \boldsymbol{p}_t)}\right), \\ &= \sum_i \Omega_i(q_t^b; \boldsymbol{p}_t)\log\left(\frac{p_{i,t+1}}{p_{it}}\right) + (1 + \Lambda_b(q_{t+1}^b; \boldsymbol{p}_{t+1}))\log\left(\frac{q_{t+1}^b}{q_t^b}\right) + O\left(\max\left\{\Delta_p^2, \Delta_q^2\right\}\right).\end{aligned}$$

Since $\widetilde{E}_b(q_{t+1}^b; \boldsymbol{p}_{t+1})$ is continuously differentiable of at least second order, we have $\Lambda_b(q_{t+1}^b; \boldsymbol{p}_{t+1}) \equiv O\left(|t-b|\Delta_p\right)$. This leads to the desired result. □

**Lemma.** *See Lemma 3.*

*Proof.* Using Lemma 8 below, we can write the Laspeyres CoL index from Equation (4) as:

$$\begin{aligned}\Lambda_b(q; \boldsymbol{p}_t) &= \frac{\partial}{\partial \log q}\log\left(\frac{\widetilde{E}_b(q; \boldsymbol{p}_t)}{\widetilde{E}_b(q; \boldsymbol{p}_b)}\right), \\ &= \frac{\partial}{\partial \log q}\int_b^t \frac{\partial}{\partial \tau}\log \widetilde{E}_b(q; \boldsymbol{p}_\tau)\, d\tau, \\ &= \frac{\partial}{\partial \log q}\int_b^t \sum_i \Omega_i(q; \boldsymbol{p}_\tau)\frac{d\log p_{i,\tau}}{d\tau}\, d\tau,\end{aligned}$$

$$= \int_b^t \sum_i \frac{\partial \Omega_i(q; \boldsymbol{p}_\tau)}{\partial \log q} \frac{d \log p_{i,\tau}}{d\tau} d\tau,$$

which leads to Equation (21). □

**Proposition.** *See Proposition 1.*

*Proof.* First, we will establish a bound on the error corresponding to the approximation of the nonhomothetic correction function $\Lambda_b(q; \boldsymbol{p}_t)$ with the nonparametric estimation $\widehat{\Lambda}^{(1)}_{t+1}(q)$. Recall from Equation (A14):

$$\Lambda_b(q; \boldsymbol{p}_t) = \frac{\partial}{\partial \log q} \log\left(\frac{\widetilde{E}_b(q; \boldsymbol{p}_{t+1})}{\widetilde{E}_b(q; \boldsymbol{p}_b)}\right) = \sum_{\tau=b}^{t} \frac{\partial}{\partial \log q} \log\left(\frac{\widetilde{E}_b(q; \boldsymbol{p}_{\tau+1})}{\widetilde{E}_b(q; \boldsymbol{p}_\tau)}\right).$$

Using Lemma 1 and in particular Equation (12), we have:

$$\log\left(\frac{\widetilde{E}_b(q^n_t; \boldsymbol{p}_{t+1})}{\widetilde{E}_b(q^n_t; \boldsymbol{p}_t)}\right) = \mathbb{P}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}\right) + O\left(\Delta_p^2\right).$$

Let us now begin with the first period away from the base period where we observe the real consumption index $q^n_b \equiv y^n_b$. We apply Lemma 10 below for $y^n \equiv \log \mathbb{P}_G\left(\boldsymbol{p}_b, \boldsymbol{p}_{b+1}; \boldsymbol{q}^n_b, \boldsymbol{q}^n_{b+1}\right)$, $x^n \equiv \log \widehat{q}^n_b$, $z^n \equiv \log q^n_b$, $v^n \equiv \delta_v \equiv 0$, and $\Delta_\varepsilon \equiv \Delta_p^2$ to find:

$$\frac{\partial}{\partial \log q} \log\left(\frac{\widetilde{E}_b(q; \boldsymbol{p}_{b+1})}{\widetilde{E}_b(q; \boldsymbol{p}_b)}\right) = \sum_{k=0}^{K_N} \widehat{\alpha}_{k,b} g'_k(q) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}} \cdot \Delta_p^4 + K_N^{-(m-1)}\right)\right),$$

where $\left(\widehat{\alpha}_{k,b}\right)_{k=0}^{K_N}$ solve Equation (24) for $t = b$. Therefore, we have:

$$\Lambda_b(q; \boldsymbol{p}_{b+1}) = \widehat{\Lambda}^{(1)}_{b+1}(q) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}} \cdot \Delta_p^4 + K_N^{-(m-1)}\right)\right).$$

Applying Lemma 2, the desired result follows for the first period.

Next, we recursively apply Lemma 10 below for $y^n \equiv \log \mathbb{P}_G\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}^n_t, \boldsymbol{q}^n_{t+1}\right)$, $x^n \equiv \log \widehat{q}^n_b$, $z^n \equiv \log q^n_b$, with $\delta_v$ denoting the error from the previous period approximation error in the ReC index and $\Delta_\varepsilon \equiv \Delta_p^2$ to find:

$$\frac{\partial}{\partial \log q} \log\left(\frac{\widetilde{E}_b(q; \boldsymbol{p}_{t+1})}{\widetilde{E}_b(q; \boldsymbol{p}_t)}\right) = \sum_{k=0}^{K_N} \widehat{\alpha}_{k,t} g'_k(q) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}} \cdot \Delta_p^4 + K_N^{-(m-1)}\right)\right),$$

where $\left(\widehat{\alpha}_{k,t}\right)_{k=0}^{K_N}$ solve Equation (24). Note that the term $O(\delta_v)$ is of the same order as the error term on the right hand side of the equation above and is therefore absorbed in that error. Therefore, we have for all $t$:

$$\Lambda_b(q;\boldsymbol{p}_t) = \widehat{\Lambda}^{(1)}_{t+1}(q) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}}\cdot\Delta_p^4 + K_N^{-(m-1)}\right)\right).$$

Applying Lemma 2, the desired result follows for all $t$. □

**Lemma 10.** *Assume that we observe $(y^n, x^n)_{n=1}^N$ such that*

$$y^n = f(z^n) + \varepsilon^n,$$
$$x^n = z^n + v^n,$$

*with $y^n, z^n, x^n \in \mathbb{R}$ where the following conditions are satisfied:*

1. *$z^n$ is distributed according to a probability distribution function that is bounded away from zero over the interval $[\underline{z}, \overline{z}]$*
2. *Function $f(\cdot)$ is continuouly differentiable of order $m$ over the interval $[\underline{z}, \overline{z}]$*
3. *$|\varepsilon^n| < \Delta_\varepsilon$ and $|v^n| < \delta_v$.*

*Let coefficients $(\widehat{\alpha}_k)_{k=0}^{K_N}$ solve the following problem:*

$$\min_{(\alpha_k)_{k=0}^{K_N}} \sum_{n=1}^N \left(y^n - \sum_{k=0}^{K_N} \alpha_k g_k(x^n)\right)^2, \tag{A16}$$

*where $g_k(q)$ is the Legendre polynomial of order $k$. Then, we have*

$$f'(q) = \sum_{k=0}^{K_N} \widehat{\alpha}_k g_k'(q) + O(\delta_v) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}}\cdot\Delta_\varepsilon^2 + K_N^{-(m-1)}\right)\right). \tag{A17}$$

*If $z^n, x^n \in \mathbb{R}^D$ for $D \geq 2$ and $z^n$ belongs to a Cartesian product of compact connected intervals, and its probability distribution is bounded away from zero over this set, Equation A17 holds for any arbitrary integer $m$ if function $f(\cdot)$ is analytical.*

*Proof.* The proof closely follows the proof of Theorem 1 of Newey (1997). Define $\boldsymbol{g}(z) \equiv \left[g_0(z), \cdots, g_{K_N}(z)\right]^t$ where superscript $t$ stands for the transpose of the matrix, and let

$$\boldsymbol{G}^* \equiv \left[\boldsymbol{g}(z^1), \cdots, \boldsymbol{g}(z^n)\right]^t,$$
$$\boldsymbol{G} \equiv \left[\boldsymbol{g}(x^1), \cdots, \boldsymbol{g}(x^n)\right]^t.$$

Define:

$$\widehat{\boldsymbol{\alpha}} \equiv (\boldsymbol{G}^t \boldsymbol{G})^{-1} \boldsymbol{G}^t \boldsymbol{y},$$
$$\boldsymbol{\alpha}^* \equiv \left((\boldsymbol{G}^*)^t \boldsymbol{G}^*\right)^{-1} (\boldsymbol{G}^*)^t \boldsymbol{y}.$$

First, assumptions (i) and (ii) correspond to Assumptions 8 and 9 in Newey (1997). This implies that Assumption 3 of Newey (1997) is satisfied for first derivative function such that:

$$\sup_{z\in[\underline{z},\overline{z}]} \left| f'(z) - \sum_{k=0}^{K_N} \alpha_k^* g_k'(z) \right| = O\left(K_N^{-(m-1)}\right).$$

Using the results of Section 5, we additionally find that

$$\xi_1(K_N) \equiv \sup_{z\in[\underline{z},\overline{z}]} \left\| \left( g_0'(z), \cdots, g_{K_N}'(z) \right) \right\| = O\left(K_N^3\right),$$

where $||\cdots||$ corresponds to the Euclidean norm. It follows from the same steps as in the proof of Theorem 1 of Newey (1997) that:

$$f'(z) = \sum_{k=0}^{K_N} \alpha_k^* g_k'(z) + O_p\left( K_N^3 \left( \sqrt{\frac{K_N}{N}} \cdot \Delta_\varepsilon^2 + K_N^{-(m-1)} \right) \right).$$

with the only difference being the fact that here $\mathbb{E}[\varepsilon_n \varepsilon_{n'}]$ is not a constant, but rather we have $\mathbb{E}[\varepsilon_n \varepsilon_{n'}] = O\left(\Delta_\varepsilon^2\right)$.

Define $\boldsymbol{g}(z) \equiv \left[ g_0'(z), \cdots, g_{K_N}'(z) \right]^t$ and note that:

$$\boldsymbol{G} = \boldsymbol{G}^* + \left[ \boldsymbol{g}'\left(x^1\right) \cdot v^1, \cdots, \boldsymbol{g}'(x^n) \cdot v^n \right]^t + O\left(\delta_v^2\right),$$

which implies:

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^* + O(\delta_v).$$

Equation (A17) then follows from the observation that:

$$\sum_{k=0}^{K_N} \alpha_k^* g_k'(z) - \sum_{k=0}^{K_N} \widehat{\alpha}_k g_k'(z) = O(\delta_v).$$

□

**Lemma.** *See Lemma 4.*

*Proof.* Once again, we start with:

$$\log\left(\frac{y_{t+1}^n}{y_t^n}\right) = \log\frac{\widetilde{E}_b\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}^n\right)}{\widetilde{E}_b\left(q_t^{b,n};\boldsymbol{p}_t^n\right)},$$

and rely on Lemma 8 for variables $\boldsymbol{x} \equiv (q;\boldsymbol{p})$ to find:

$$\begin{aligned}
\log\left(\frac{y_{t+1}^n}{y_t^n}\right) &= \frac{1}{2}\sum_{i=1}^{I}\left[\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p})}{\partial \log p_i}\right|_{(q;\boldsymbol{p})\equiv\left(q_t^{b,n};\boldsymbol{p}_t^n\right)} + \left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p})}{\partial \log p_i}\right|_{(q;\boldsymbol{p})\equiv\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}^n\right)}\right]\cdot\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right) \\
&\quad + \frac{1}{2}\left[\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p})}{\partial \log q}\right|_{(q;\boldsymbol{p})\equiv\left(q_t^{b,n};\boldsymbol{p}_t^n\right)} + \left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p})}{\partial \log q}\right|_{(q;\boldsymbol{p})\equiv\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}^n\right)}\right]\cdot\log\left(\frac{q_{t+1}^{b,n}}{q_t^{b,n}}\right) \\
&\quad + O\left(\Delta^3\right), \qquad \Delta \equiv \max\left\{\Delta_p,\Delta_q\right\}, \\
&= \frac{1}{2}\sum_{i=1}^{I}\left[\Omega_i^b\left(q_t^{b,n};\boldsymbol{p}_t\right) + \Omega_i^b\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}\right)\right]\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right) \\
&\quad + \left(1 + \frac{1}{2}\left[\Lambda_b\left(q_t^{b,n};\boldsymbol{p}_t\right) + \Lambda_b\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}\right)\right]\right)\cdot\log\left(\frac{q_{t+1}^{b,n}}{q_t^{b,n}}\right) + O\left(\Delta^3\right),
\end{aligned}$$

where in the second equality we have again used Shephard's lemma, as well as the definition of the first-order nonhomotheticity correction function. □

**Proposition.** *See Proposition 2.*

*Proof.* As in the proof of Proposition 1, we will first establish a bound on the error corresponding to the approximation of the nonhomothetic correction function $\Lambda_b(q;\boldsymbol{p}_t)$ with the nonparametric estimation $\widehat{\Lambda}_{t+1}^{(2)}(q)$. Recall from Equation (A14):

$$\Lambda_b(q;\boldsymbol{p}_t) = \frac{\partial}{\partial \log q}\log\left(\frac{\widetilde{E}_b(q;\boldsymbol{p}_{t+1})}{\widetilde{E}_b(q;\boldsymbol{p}_b)}\right) = \sum_{\tau=b}^{t}\frac{\partial}{\partial \log q}\log\left(\frac{\widetilde{E}_b(q;\boldsymbol{p}_{\tau+1})}{\widetilde{E}_b(q;\boldsymbol{p}_\tau)}\right).$$

To approximate this function, we first note that

$$\begin{aligned}
\log\left(\frac{\widetilde{E}_b\left(q_t^{b,n};\boldsymbol{p}_{t+1}\right)}{\widetilde{E}_b\left(q_t^{b,n};\boldsymbol{p}_t\right)}\right) &= \frac{1}{2}\sum_{i=1}^{I}\left[\Omega_i^b\left(q_t^{b,n};\boldsymbol{p}_t\right) + \Omega_i^b\left(q_t^{b,n};\boldsymbol{p}_{t+1}\right)\right]\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right), \\
&= \mathbb{P}_T\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{q}_{t+1}^n\right)
\end{aligned}$$

$$-\frac{1}{4}\sum_{i=1}^{I}\left(\frac{\partial\Omega_i^b\left(q_t^{b,t};\boldsymbol{p}_{t+1}\right)}{\partial\log q}+\frac{\partial\Omega_i^b\left(q_t^{b,t};\boldsymbol{p}_{t+1}\right)}{\partial\log q}\right)\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right), \tag{A18}$$

where we have used

$$\Omega_i^b\left(q_t^{b,n};\boldsymbol{p}_{t+1}\right)=\Omega_i^b\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}\right)-\frac{1}{2}\left(\frac{\partial\Omega_i^b\left(q_t^{b,t};\boldsymbol{p}_{t+1}\right)}{\partial\log q}+\frac{\partial\Omega_i^b\left(q_t^{b,t};\boldsymbol{p}_{t+1}\right)}{\partial\log q}\right)\log\left(\frac{q_{t+1}^{b,n}}{q_t^{b,n}}\right)+O\left(\Delta_q^3\right).$$

Defining

$$\mathscr{P}_1^{\dagger}\left(q^b;\boldsymbol{p}_t,\boldsymbol{p}_{t+1}\right)\equiv\frac{\partial}{\partial\log q}\left[\sum_{i=1}^{I}\Omega_i^b\left(q^b;\boldsymbol{p}_{t+1}\right)\cdot\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)\right],$$

we can now rewrite Equation (A18) as:

$$\log\left(\frac{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_{t+1}\right)}{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_t\right)}\right)=\mathbb{P}_T\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{q}_{t+1}^n\right)-\frac{1}{4}\left[\mathscr{P}_1^{\dagger}\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{p}_{t+1}\right)+\mathscr{P}_1^{\dagger}\left(q_{t+1}^n;\boldsymbol{p}_t,\boldsymbol{p}_{t+1}\right)\right]\log\left(\frac{q_{t+1}^n}{q_t^n}\right)+O\left(\Delta_p^3\right). \tag{A19}$$

The key observation is to note that, through the definition of the geometric index, we have:

$$\sum_{i=1}^{I}\Omega_i^b\left(q_{t+1}^{b,n};\boldsymbol{p}_{t+1}\right)\cdot\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)=-\log\mathbb{P}_G\left(\boldsymbol{p}_{t+1},\boldsymbol{p}_t;\boldsymbol{q}_{t+1}^n,\boldsymbol{q}_t^n\right).$$

We now apply Lemma 10 for $y^n\equiv\log\mathbb{P}_G\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{q}_{t+1}^n\right)$, $x^n\equiv\log\widehat{q}_b^n$, $z^n\equiv\log q_b^n$, $v^n\equiv\delta_v\equiv 0$, and $\Delta_\varepsilon\equiv\Delta_p^2$ to find that for $\lambda_t^n$ defined by Equations (34) and (37), we have:

$$-\frac{1}{4}\left[\mathscr{P}_1^{\dagger}\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{p}_{t+1}\right)+\mathscr{P}_1^{\dagger}\left(q_{t+1}^n;\boldsymbol{p}_t,\boldsymbol{p}_{t+1}\right)\right]=\lambda_t^n+O\left(\epsilon\right)+O_p\left(K_N^{4-m}\right).$$

Therefore, we can now rewrite Equation (A19) as:

$$\log\left(\frac{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_{t+1}\right)}{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_t\right)}\right)=\mathbb{P}_T\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{q}_{t+1}^n\right)+\lambda_t^n+O\left(\epsilon\right)+O_p\left(K_N^{4-m}\right)+O\left(\Delta_p^3\right). \tag{A20}$$

This allows us to again apply Lemma 10 for $y^n\equiv\log\mathbb{P}_T\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{q}_{t+1}^n\right)+\lambda_t^n$, $x^n\equiv\log\widehat{q}_b^n$, $z^n\equiv\log q_b^n$, and $\Delta_\varepsilon\equiv O\left(\epsilon\right)+O_p\left(K_N^{4-m}\right)+O\left(\Delta_p^3\right)$ recursively (similar to that in the proof of Proposition 1) to find Equation (40). □

**Lemma.** *See Lemma 5.*

*Proof.* First, note that we have:

$$\log\left(\frac{y_{t+1}^n}{y_t^n}\right)=\log\frac{\widetilde{E}_b\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)}{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)},$$

We can do a first-order Taylor expansion of the left-hand-side of the equation above in terms of $q_{t+1}^n$, assuming that $\log\left(\frac{q_{t+1}^n}{q_t^n}\right)<\Delta_q$:

$$\log\left(\frac{y_{t+1}^n}{y_t^n}\right)=\log\frac{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)}+\sum_{d=1}^{D}\left.\frac{\partial\log\widetilde{E}_b\left(q;\boldsymbol{p},\boldsymbol{\theta}\right)}{\partial\log\theta_d}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv\left(q_t^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}\cdot\log\left(\frac{\theta_{d,t+1}^n}{\theta_{d,t}^n}\right)$$
$$+\left.\frac{\partial\log\widetilde{E}_b\left(q;\boldsymbol{p},\boldsymbol{\theta}\right)}{\partial\log q}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv\left(q_t^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}\cdot\log\left(\frac{q_{t+1}^n}{q_t^n}\right)+O\left(\max\left\{\Delta_q,\Delta_\theta\right\}^2\right).$$

This allows us to write:

$$\log\left(\frac{q_{t+1}^n}{q_t^n}\right)=\frac{\log\left(\frac{y_{t+1}^n}{y_t^n}\right)-\sum_{d=1}^{D}\Gamma_{b,d}^{(1)}\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)\cdot\log\left(\frac{\theta_{d,t+1}^n}{\theta_{d,t}^n}\right)}{1+\Lambda_b^{(1)}\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)},$$

where we have defined:

$$\begin{aligned}\Lambda_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)&=\frac{\partial\log\widetilde{E}_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}{\partial\log q}-\frac{\partial\log\widetilde{E}_b\left(q;\boldsymbol{p}_b,\boldsymbol{\theta}_t^n\right)}{\partial\log q},\\&=\frac{\partial}{\partial\log q}\log\left(\frac{\widetilde{E}_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}{\widetilde{E}_b\left(q;\boldsymbol{p}_b,\boldsymbol{\theta}_t^n\right)}\right),\end{aligned}\tag{A21}$$

noting that $\frac{\partial\log\widetilde{E}_b(q;\boldsymbol{p}_b,\boldsymbol{\theta}_t^n)}{\partial\log q}\equiv 1$ for all $\boldsymbol{\theta}_t^n$, and:

$$\begin{aligned}\Gamma_{b,d}\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)&=\frac{\partial\log\widetilde{E}_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}{\partial\log\theta_d}-\frac{\partial\log\widetilde{E}_b\left(q;\boldsymbol{p}_b,\boldsymbol{\theta}_n^n\right)}{\partial\log\theta_d},\\&=\frac{\partial}{\partial\log\theta_d}\log\left(\frac{\widetilde{E}_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}{\widetilde{E}_b\left(q;\boldsymbol{p}_b,\boldsymbol{\theta}_t^n\right)}\right),\end{aligned}\tag{A22}$$

noting that $\frac{\partial\log\widetilde{E}_b(q;\boldsymbol{p}_b,\boldsymbol{\theta}_t^n)}{\partial\log\theta_d}\equiv 0$ for all $\boldsymbol{\theta}_t^n$. □

**Lemma 11.** *If Assumption (2) holds at some period $t$, then for any well-behaved underlying utility* *function $U$, the Laspeyres CoL index of any household $n$ is related to the geometric and Törqvist price indices for that household to the first order of approximation through*

$$\log \mathscr{P}_{CL}\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{\theta}_t^n\right)=\log \mathbb{P}_G\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{q}_{t+1}^n\right)+O\left(\Delta_p^2\right). \tag{A23}$$

*Proof.* Performing of a Taylor series expansion, we find:

$$\begin{aligned}\log \mathscr{P}_{CL}\left(\boldsymbol{p}_t,\boldsymbol{p}_{t+1};\boldsymbol{q}_t^n,\boldsymbol{\theta}_t^n\right)&=\log \frac{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_t^n\right)}{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)},\\ &=\sum_{i=1}^{I}\Omega_i\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)+O\left(\Delta_p^2\right).\end{aligned} \tag{A24}$$

From Shephard's lemma in Equation (17), we have $\Omega_i\left(q_t^n;\boldsymbol{p}_t\right)=s_{it}^n$, leading to Equation (A23). □

**Lemma.** *See Lemma 4.*

*Proof.* Once again, we start with:

$$\log\left(\frac{y_{t+1}^n}{y_t^n}\right)=\log \frac{\widetilde{E}_b\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)}{\widetilde{E}_b\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)},$$

and rely on Lemma 8 for variables $\boldsymbol{x}\equiv(q;\boldsymbol{p})$ to find:

$$\begin{aligned}\log\left(\frac{y_{t+1}^n}{y_t^n}\right)=&\frac{1}{2}\sum_{i=1}^{I}\left[\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p},\boldsymbol{\theta})}{\partial \log p_i}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n)}+\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p},\boldsymbol{\theta})}{\partial \log p_i}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)}\right]\cdot\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)\\ &+\frac{1}{2}\sum_{d=1}^{D}\left[\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p},\boldsymbol{\theta})}{\partial \log \theta_d}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n)}+\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p},\boldsymbol{\theta})}{\partial \log \theta_d}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)}\right]\cdot\log\left(\frac{\theta_{d,t+1}}{\theta_{d,t}}\right)\\ &+\frac{1}{2}\left[\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p},\boldsymbol{\theta})}{\partial \log q}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n)}+\left.\frac{\partial \log \widetilde{E}_b(q;\boldsymbol{p},\boldsymbol{\theta})}{\partial \log q}\right|_{(q;\boldsymbol{p},\boldsymbol{\theta})\equiv\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)}\right]\cdot\log\left(\frac{q_{t+1}^n}{q_t^n}\right)\\ &+O\left(\Delta^3\right),\qquad \Delta\equiv\max\left\{\Delta_p,\Delta_q,\Delta_\theta\right\},\\ =&\frac{1}{2}\sum_{i=1}^{I}\left[\Omega_i\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)+\Omega_i\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)\right]\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)\\ &+\frac{1}{2}\sum_{d=1}^{D}\left[\Gamma_{b,d}\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)+\Gamma_{b,d}\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)\right]\cdot\log\left(\frac{\theta_{d,t+1}}{\theta_{d,t}}\right)\\ &+\left(1+\frac{1}{2}\left[\Lambda_b\left(q_t^n;\boldsymbol{p}_t,\boldsymbol{\theta}_t^n\right)+\Lambda_b\left(q_{t+1}^n;\boldsymbol{p}_{t+1},\boldsymbol{\theta}_{t+1}^n\right)\right]\right)\cdot\log\left(\frac{q_{t+1}^n}{q_t^n}\right)+O\left(\Delta^3\right),\end{aligned}$$

where in the second equality we have again used Shephard's lemma, as well as the definition of the first-order nonhomotheticity correction function. □

**Proposition.** *See Proposition 3.*

*Proof.* We need to establish bounds on the error corresponding to the approximations of the correction functions $\Lambda_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)$ and $\Gamma_{b,d}\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)$and with the by $\widehat{\Lambda}^{(1)}_{t+1}(q,\boldsymbol{\theta})$ and $\widehat{\Gamma}^{(1)}_{d,t+1}(q,\boldsymbol{\theta})$. The former follows closely that in the proof of Proposition 1, except that we now invoke the multi-dimensional case of Lemma 10, requiring the CoL index to be a infinitely differentiable. This leads us to:

$$\Lambda_b\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)=\widehat{\Lambda}^{(1)}_{t+1}(q,\boldsymbol{\theta})+O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}}\cdot\Delta_p^4+K_N^{-m}\right)\right),$$

$$\Gamma_{b,d}\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)=\widehat{\Gamma}^{(1)}_{d,t+1}(q,\boldsymbol{\theta})+O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}}\cdot\Delta_p^4+K_N^{-m}\right)\right),$$

where $m$ is any positive number. Applying Lemma 5, the desired result follows. □

**Algorithm 4.** *Let $\widehat{q}^n_b\equiv q^n_b\equiv y^n_b$ and define function $\widehat{\mathscr{P}}^{(2)}_b(\cdot,\cdot)$ such that $\widehat{\mathscr{P}}^{(2)}_b(q,\boldsymbol{\theta})\equiv 1$, consider a sequence $\{g_k(q,\boldsymbol{\theta})\}_{k=0}^{K_N}$ of power functions of $q$ and $\boldsymbol{\theta}$ where $N$ is the number of households in the cross-section. For each $t\geq b$, apply the following steps:*

1. *Initialize the values of the real consumption $\widehat{q}^{n,(0)}_{t+1}$ for each household at $t+1$ using Equations (23)–(26) as in Algorithm 1.*

2. *For each $t\geq b$, iterate over the following steps over $\tau\in\{0,1,\cdots\}$ until convergance for some $\epsilon\ll 1$:*

   (a) *Solve for the coefficients $\left(\widehat{\alpha}^{\dagger}_{k,t}\right)_{k=0}^{K}$ in the following problem:*

$$\min_{\left(\alpha^{\dagger}_{k,t}\right)_{k=0}^{K}}\sum_{n=1}^{N}\left(\log\mathbb{P}_G\left(\boldsymbol{p}_{t+1},\boldsymbol{p}_t;\boldsymbol{q}^n_{t+1},\boldsymbol{q}^n_t\right)-\sum_{k=0}^{K}\alpha^{\dagger}_{k,t}\,g_k\left(\widehat{q}^{n,(\tau)}_{t+1},\boldsymbol{\theta}^n_{t+1}\right)\right)^2. \tag{A25}$$

   (b) *Update the next period function $\widehat{\mathscr{P}}^{(2)}_{t+1}(\cdot,\cdot)$:*

$$\log\widehat{\mathscr{P}}^{(2)}_{t+1}(q,\boldsymbol{\theta})\equiv\log\widehat{\mathscr{P}}^{(2)}_t(q,\boldsymbol{\theta})+\sum_{k=0}^{K}\widehat{\beta}_{k,t}\,g_k(q,\boldsymbol{\theta}), \tag{A26}$$

*where the coefficients* $\left(\widehat{\beta}_{k,t}\right)_{k=0}^{K}$ *solve the following problem:*

$$\min_{\left(\widehat{\alpha}_{k,t}\right)_{k=0}^{K}} \sum_{n=1}^{N} \left( \log \mathbb{P}_T\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right) + \lambda_t^{n,(\tau)} - \sum_{k=0}^{K} \widehat{\beta}_{k,t}\, g_k\left(\widehat{q}_t^n, \boldsymbol{\theta}_t^n\right)\right)^2, \tag{A27}$$

*with* $\lambda_t^{n,(\tau)}$ *is defined as:*

$$\lambda_t^{n,(\tau)} \equiv \frac{1}{4}\sum_{k=0}^{K} \widehat{\alpha}_{k,t}^{\dagger} \left[ \frac{\partial g_k\left(\widehat{q}_t^n, \boldsymbol{\theta}_t^n\right)}{\partial \log q} + \frac{\partial g_k\left(\widehat{q}_{t+1}^{n,(\tau)}, \boldsymbol{\theta}_{t+1}^n\right)}{\partial \log q} \right] \log\left(\frac{\widehat{q}_{t+1}^{n,(\tau)}}{\widehat{q}_t^n}\right) \tag{A28}$$

$$+ \frac{1}{4}\sum_{d=1}^{D}\sum_{k=0}^{K} \widehat{\alpha}_{k,t}^{\dagger} \left[ \frac{\partial g_k\left(\widehat{q}_t^n, \boldsymbol{\theta}_t^n\right)}{\partial \log \theta_d} + \frac{\partial g_k\left(\widehat{q}_{t+1}^{n,(\tau)}, \boldsymbol{\theta}_{t+1}^n\right)}{\partial \log \theta_d} \right] \log\left(\frac{\theta_{d,t+1}}{\theta_{d,t}}\right). \tag{A29}$$

(c) *Update the real consumption in the next period for each household:*

$$\log \widehat{q}_{t+1}^{n,(\tau+1)} = \log \widehat{q}_t^{n,(\tau)} + \frac{1}{1 + \frac{1}{2}\left[\widehat{\Lambda}_t^{(2)}\left(\widehat{q}_t^n, \boldsymbol{\theta}_t^n\right) + \widehat{\Lambda}_{t+1}^{(2)}\left(\widehat{q}_{t+1}^{n,(\tau)}, \boldsymbol{\theta}_{t+1}^n\right)\right]} \tag{A30}$$

$$\times \left[ \log\left(\frac{y_{t+1}^n / y_t^n}{\mathbb{P}_T\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right)}\right) - \log \widehat{\mathbb{C}}^{(2)}\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n\right)\right], \tag{A31}$$

*where we have defined the second-order covariate index*

$$\widehat{\mathbb{C}}^{(2)}\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n, \boldsymbol{\theta}_{t+1}^n\right) \equiv \frac{1}{2}\sum_{d=1}^{D}\left[\widehat{\Gamma}_{d,t}\left(q_t^n, \boldsymbol{\theta}_t^n\right) + \widehat{\Gamma}_{d,t+1}\left(q_{t+1}^n, \boldsymbol{\theta}_{t+1}^n\right)\right] \cdot \log\left(\frac{\theta_{d,t+1}}{\theta_{d,t}}\right)$$

*and the approximate correction functions as:*

$$\widehat{\Lambda}_{t+1}^{(2)}(q, \boldsymbol{\theta}) \equiv \frac{\partial}{\partial \log q} \log \widehat{\mathscr{P}}_{t+1}^{(2)}(q, \boldsymbol{\theta}) = \sum_{k=0}^{K}\left(\sum_{\tau=b+1}^{t+1} \widehat{\beta}_{k,\tau}\right)\frac{\partial g_k(q, \boldsymbol{\theta})}{\partial \log q}, \tag{A32}$$

$$\widehat{\Gamma}_{d,t}^{(2)}(q, \boldsymbol{\theta}) \equiv \frac{d}{d \log \theta_d} \log \widehat{\mathscr{P}}_{t+1}^{(1)}(q, \boldsymbol{\theta}) = \sum_{k=0}^{K_N}\left(\sum_{\tau=b+1}^{t+1} \widehat{\alpha}_{k,\tau}\right)\frac{\partial g_k(q, \boldsymbol{\theta})}{\partial \log \theta_d}, \tag{A33}$$

(d) *Stop if* $\max_n \left|\widehat{q}_{t+1}^{n,(\tau+1)} - \widehat{q}_{t+1}^{n,(\tau)}\right| < \epsilon$ *and set* $\widehat{q}_{t+1}^n \equiv \widehat{q}_{t+1}^{n,(\tau+1)}$.

**Proposition 4.** *Assume that the expenditure function* $\log \widetilde{E}_b(\cdot;\cdot)$ *is an analytical function. If Assumptions 2, 4, and 3 hold, then the sequences of real consumptions* $\boldsymbol{q}_{(b,T)}^n$ *constructed by Algorithm*

*4 satisfy:*

$$\log \mathcal{Q}_{RC}\left(\boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n; \boldsymbol{p}_b\right) = \log\left(\frac{\widehat{q}_{t+1}^n}{\widehat{q}_t^n}\right) + O\left(\Delta^2\right) + O_p\left(K_N^3\left(\sqrt{\frac{K_N}{N}} \cdot \Delta^4 + K_N^{-m}\right)\Delta\right), \tag{A34}$$

*where* $\Delta \equiv \max\left\{\Delta_p, \Delta_q, \Delta_\theta\right\}$ *and* $m$ *is any positive number.*

*Proof.* We need to first find an approximation to $\mathscr{P}_{CL}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{\theta}_t^n\right)$. Applying Lemma 8, we have

$$\begin{aligned}
\log \mathscr{P}_{CL}\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{\theta}_t^n\right) &= \log \frac{\widetilde{E}_b\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right)}{\widetilde{E}_b\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{\theta}_t^n\right)}, \\
&= \frac{1}{2}\sum_{i=1}^{I}\left[\Omega_i\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{\theta}_t^n\right) + \Omega_i\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right)\right] \log\left(\frac{p_{i,t+1}}{p_{i,t}}\right) + O\left(\Delta_p^3\right).
\end{aligned}$$

For the second term inside the square bracket above, using Lemma 8 again we have:

$$\begin{aligned}
\Omega_i\left(q_{t+1}^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^n\right) = \Omega_i\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right) &+ \frac{1}{2}\left(\left.\frac{\partial \Omega_i\left(q; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right)}{\partial \log q}\right|_{q=q_t^n} + \left.\frac{\partial \Omega_i\left(q; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^n\right)}{\partial \log q}\right|_{q=q_{t+1}^n}\right) \cdot \log\left(\frac{q_{t+1}^n}{q_t^n}\right) \\
&+ \frac{1}{2}\sum_{d=1}^{D}\left(\left.\frac{\partial \Omega_i\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}\right)}{\partial \log \theta_d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^n} + \left.\frac{\partial \Omega_i\left(q_{t+1}^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}\right)}{\partial \log \theta_d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{t+1}^n}\right) \cdot \log\left(\frac{\theta_{d,t+1}^n}{\theta_{d,t}^n}\right) \\
&+ O\left(\max\{\Delta_q^3, \Delta_\theta^3\}\right).
\end{aligned}$$

Thus, we find:we have:

$$\log\left(\frac{\widetilde{E}_b\left(q_t^n; \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right)}{\widetilde{E}_b\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{\theta}_t^n\right)}\right) = \mathbb{P}_T\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}; \boldsymbol{q}_t^n, \boldsymbol{q}_{t+1}^n\right) \tag{A35}$$

$$-\frac{1}{4}\left[\mathscr{P}_q^\dagger\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right) + \mathscr{P}_q^\dagger\left(q_{t+1}^n; \boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^n\right)\right] \cdot \log\left(\frac{q_{t+1}^n}{q_t^n}\right) \tag{A36}$$

$$-\frac{1}{4}\sum_{d=1}^{D}\left[\mathscr{P}_{\theta_d}^\dagger\left(q_t^n; \boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_t^n\right) + \mathscr{P}_{\theta_d}^\dagger\left(q_{t+1}^n; \boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^n\right)\right] \cdot \log\left(\frac{\theta_{d,t+1}^n}{\theta_{d,t}^n}\right) + O\left(\max\right. \tag{A37}$$

where we have defined:

$$\mathscr{P}_q^{\dagger}\left(q;\boldsymbol{p}_t,\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)\equiv\frac{\partial}{\partial\log q}\left[\sum_{i=1}^{I}\Omega_i\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)\cdot\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)\right],$$
$$\mathscr{P}_{\theta_d}^{\dagger}\left(q;\boldsymbol{p}_t,\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)\equiv\frac{\partial}{\partial\log\theta_d}\left[\sum_{i=1}^{I}\Omega_i\left(q;\boldsymbol{p}_{t+1},\boldsymbol{\theta}\right)\cdot\log\left(\frac{p_{i,t+1}}{p_{i,t}}\right)\right].$$

The remainder of the proof follows along the structure of the proof of Proposition 2. □

# B Data Appendix

*CPI dataset.* The majority of our datasets are linked to the Consumer Price Index (CPI) data series, which contain monthly or quarterly price indexes for over 200 detailed product categories, with various time frames for availability.[37] To obtain a balanced panel of inflation series derived from the CPI price indexes, whenever a category is missing we use a more aggregate series in the product hierarchy as proxy, since higher-level series usually has longer time coverage. If broad categories also have limited data availability, but one or more immediate sub-categories are populated, then we take the simple average of sub-categories as a proxy.

*Main linked CEX-CPI dataset.* The preferred dataset contains 19 expenditure product categories that collectively cover the full consumption basket from 1953 to 2019.[38] These product categories are defined in the annual summary tables of expenditures from the Consumer Expenditure Survey (CEX) published by the U.S. Bureau of Labor Statistics, which include expenditure shares across all items by quintiles of income before taxes from 1984 to 2019. For prior year, we extrapolate expenditure shares based on the available data in 1984. Using a crosswalk we build by hand based on matching of product description and aggregation level, these 19 categories are each mapped to one or more inflation series from the CPI price data. Table 1 below reports a sample of the crosswalk we build to link price and expenditure categories.

*Robustness dataset #1.* The first alternative dataset for robustness check uses the official consumption weights used by the Bureau of Labor Statistics for calculating the U.S. overall price index.[39] This dataset is limited in terms of the numbers of product categories but has the benefit of an extended time frame. The ten broad product categories included in this dataset are: food and

[37] The data is available at https://download.bls.gov/pub/time.series/cu.

[38] The 19 categories are: alcoholic beverages; apparel and services; entertainement; food at home; food away from home; gasoline, other fuels, and motor oil; healthcare; households furnishings and equipment; household operations; housekeeping supplies; miscellaneous; public and other transportation; vehicle purchases; other vehicle expenses; personal care products and services; reading; shelter; tobacco products and smoking supplies; utilities, fuels, and public services.

[39] The official consumption weights are available at https://www.bls.gov/cpi/tables/relative-importance/home.htm.

## Table A1: Crosswalk Between CEX Product Category and CPI Price Series

| Product Category in CEX Data | Linked CPI Series |
|---|---|
| Alcoholic beverages | CUSR0000SAF116 – Alcoholic beverages |
| Apparel and services | CUSR0000SAA – Apparel |
| | CUUR0000SEGD03 – Laundry and dry cleaning services |
| | CUUR0000SEGD04 – Apparel services other than laundry and dry cleaning |
| Entertainment | CUUR0000SERA – Video and audio; CUSR0000SERC – Sporting goods |
| | CUUR0000SERD – Photography; CUSR0000SERE – Other recreational goods |
| | CUSR0000SERF – Other recreation services |
| | CUUR0000SEEE03 – Internet services and electronic information providers |
| Food at home | CUSR0000SAF11 – Food at home |
| Food away from home | CUSR0000SEFV – Food away from home |
| Gasoline, other fuels, and motor oil | CUSR0000SETB – Motor fuel; CUUR0000SS47021 – Motor oil, coolant, and fluids |
| Healthcare | CUSR0000SAM – Medical care |
| Household furnishings and equipment | CUUR0000SEHH – Window and floor coverings and other linens |
| | CUSR0000SEHJ – Furniture and bedding |
| | CUUR0000SEHK – Appliances |
| | CUUR0000SEHL – Other household equipment and furnishings |
| | CUUR0000SEHM – Tools, hardware, outdoor equipment and supplies |
| | CUUR0000SEEE01 – Computers, peripherals, and smart home assistants |
| Household operations | CUUR0000SEHP – Household operations |
| Housekeeping supplies | CUSR0000SEHN – Housekeeping supplies; CUUR0000SEEC01 – Postage |
| Miscellaneous | CUUR0000SEGD – Miscellaneous personal services; CUUR0000SEGE – Miscellaneous personal goods |
| Public and other transportation | CUSR0000SETG – Public transportation |
| Vehicle purchases | CUSR0000SETA – New and used motor vehicles |
| Other vehicle expenses | CUUR0000SETC – Motor vehicle parts and equipment |
| | CUSR0000SETD – Motor vehicle maintenance and repair |
| | CUSR0000SETE – Motor vehicle insurance; CUUR0000SETF – Motor vehicle fees |
| Personal care products and services | CUUR0000SEGB – Personal care products; CUUR0000SEGC – Personal care services |
| Reading | CUSR0000SERG – Recreational reading materials |
| Shelter | CUSR0000SAH1 – Shelter |
| Tobacco products and smoking supplies | CUSR0000SEGA – Tobacco and smoking products |
| Utilities, fuels, and public services | CUSR0000SAH2 – Fuels and utilities; CUUR0000SEED – Telephone services |

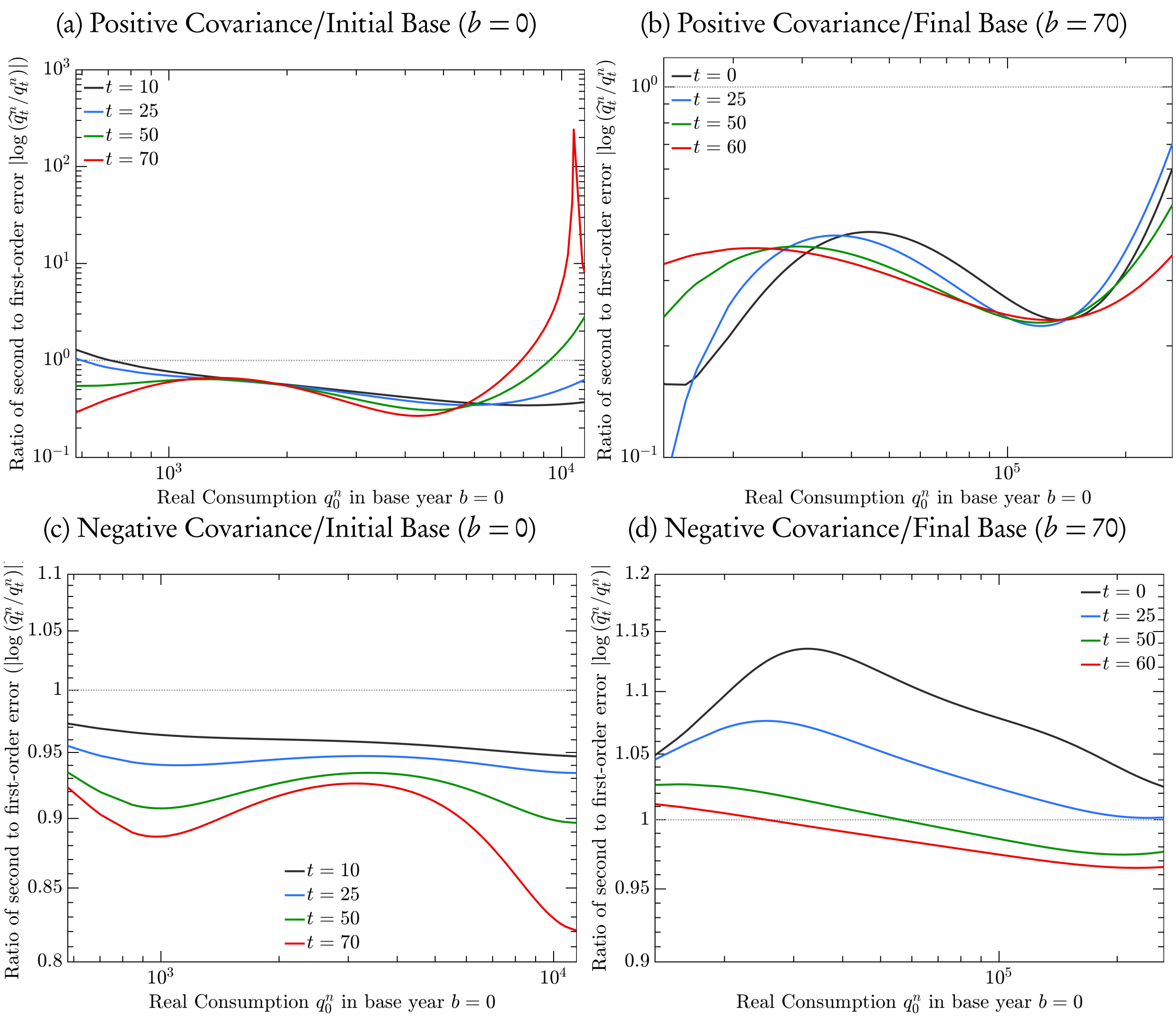
beverages, shelter, fuels and utilities, household furnishings and operations, apparel, transportation, medical care, recreation, education and communication, other goods and services. Due to the evolution of product categories and product hierarchy over the years, some sub-categories are reassigned by BLS from one broad category to another over time. For example, BLS places "Telephone services" under housing until 1997, then under "Education and communication." To address this issue, we adjust the placement of certain sub-categories and their allocated weights so that the composition of broad categories remains consistent from 1953 to 2020. In addition to the aggregate consumption weights, our linked dataset uses aggregate expenditures across income quintiles from the Consumer Expenditure Survey summary tables published by the BLS, which is available from 1984 onwards, as in the main dataset. Prior to 1984, we assume the expenditure levels to be constant and identical to 1984. We use the expenditure shares of each income quintile to distribute aggregate consumption across income groups, so that we obtain a linked dataset with consumption patterns by income groups while keeping aggregate consumption weights identical to the official weights of the BLS.

*Data on consumption growth by income groups.* The CEX annual summary tables, which we use to obtain expenditures and expenditure share by income groups, also contain information on total average annual expenditure as well as average income before and after tax for each quintile. This data is used to measure the nominal growth rates of consumption by income quintiles from 1984 to 2000. Prior to 1984, we assign a common growth rate to all income groups using the growth rate of per capita personal consumption expenditures, as measured by the BEA.

*Nielsen data.* As a third robustness exercise, we study product-level data covering consumer packaged goods. The Nielsen Homescan Consumer Panel (HMS) records consumption from 2004 to 2015 for a rotating panel of about 50,000 households, who are instructed to scan any product they purchase that has a barcode. These products are typically found in department stores, grocery stores, drug stores, convenience stores, and other similar retail outlets. Price and expenditure data are observed for each barcode, along with the socio-demographic characteristics of the participants, including household income.

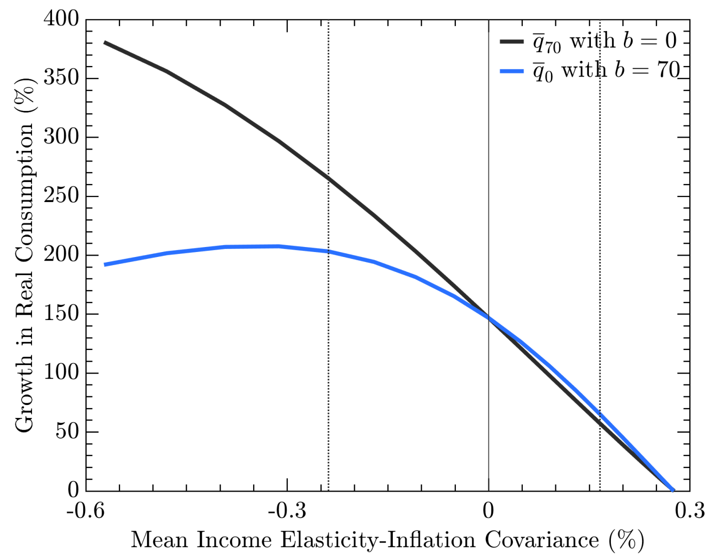
# C Additional Results

Figure A1: nhCES Example: Second vs. First-order Correction




*Note:* The figures compare the error in the approximate value of real consumption between the the first-order and second-order algorithms. The correct value of real consumption is calculated based on the underlying parameters of the nhCES preferences. The panels show the error for the choices of base period (a) $b = 0$ and (b) $b = 70$ with the positive income elasticity-inflation covariance and (c) $b = 0$ and (d) $b = 70$ with the negative covariance.

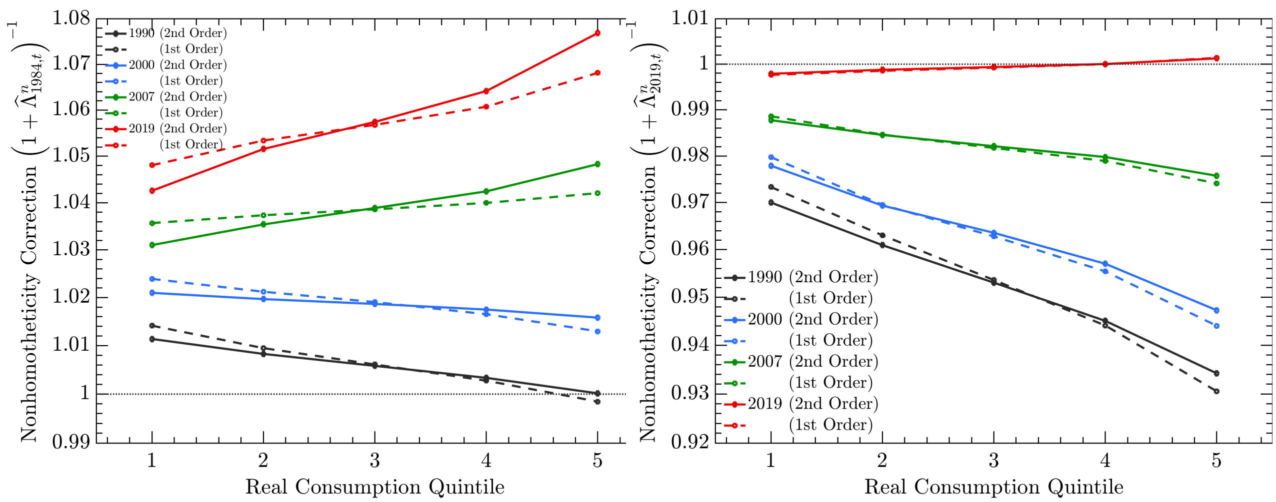
Figure A2: Example: Real Consumption Growth and Income Elasticity-Inflation Covariance




*Note:* The figure compares the growth in average real consumption for the initial and final periods as base, respectively, as a function of the mean covariance between price inflations and expenditure elasticities over the period.

Figure A3: CEX and BLS Data: Nonhomotheticity Correction

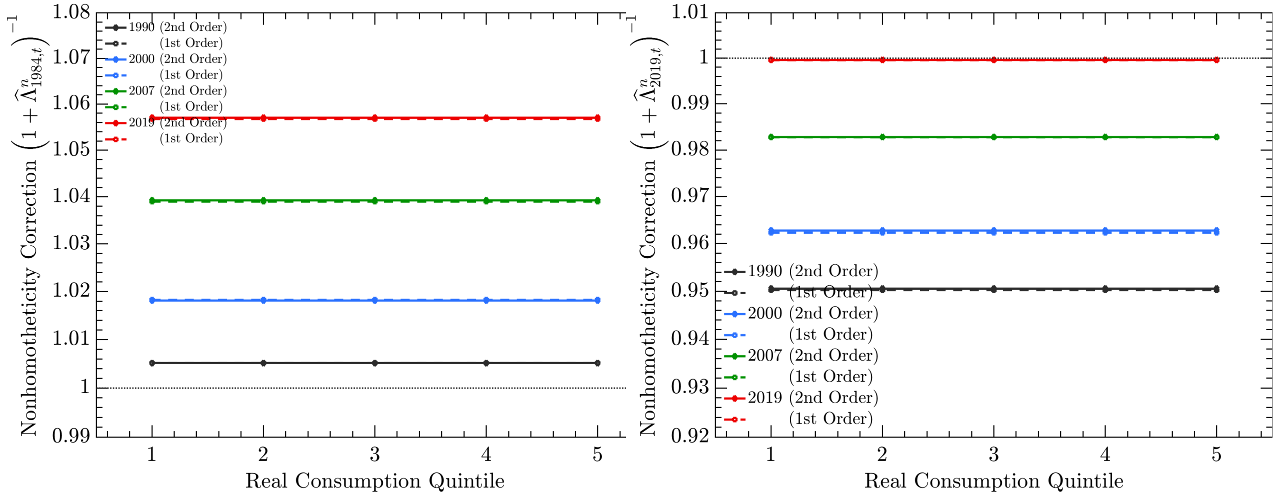
(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the "Nonhomotheticity Correction" varies over time for different quintiles of income for the initial and final years as the base period.

Figure A4: CEX and BLS Data: Nonhomotheticity Correction ($K = 1$)

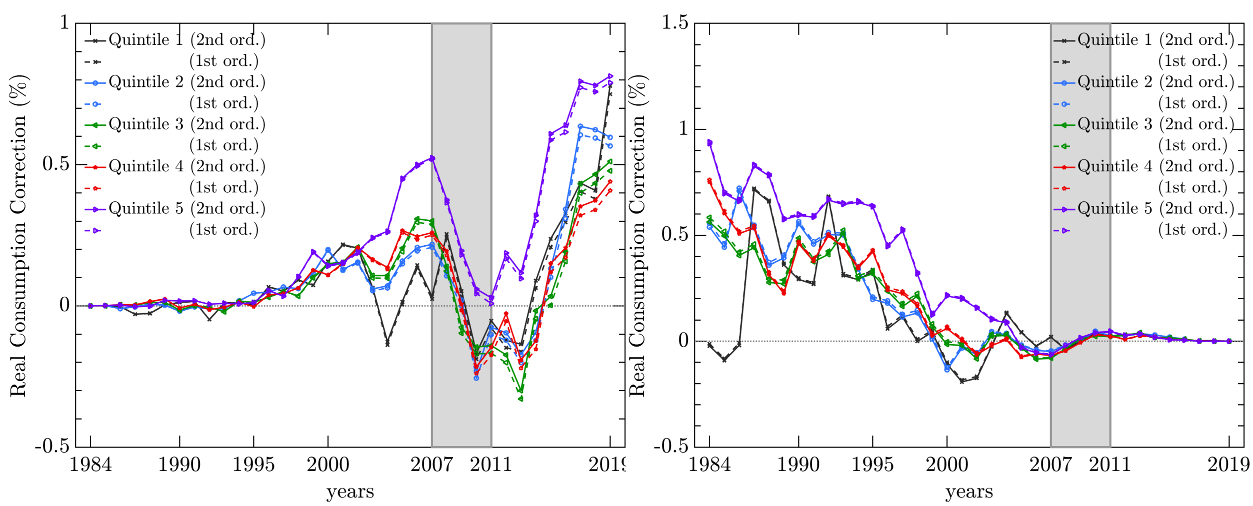
(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the "Nonhomotheticity Correction" varies over time for different quintiles of income for the initial and final years as the base period with $K = 1$.

Figure A5: CEX and BLS Data: Corrected relative to Uncorrected Real Consumption ($K = 1$)

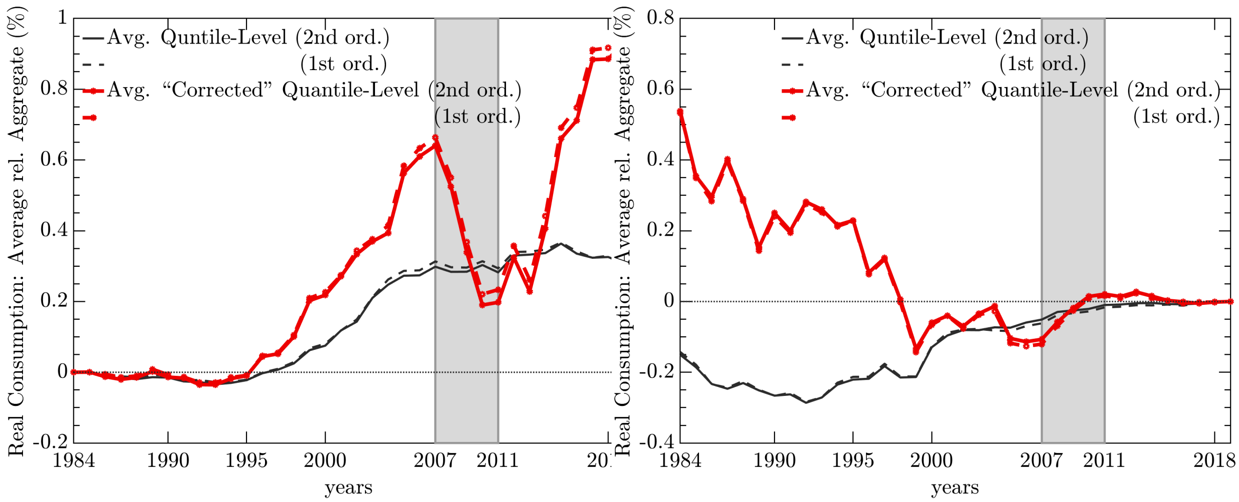
(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the ratio of the corrected to the uncorrected measures of real consumption vary over time for each quintile of initial real consumption for the initial and final years as the base period, respectively, with $K = 1$. The great recession has been indicated in grey background.

## Figure A6: CEX and BLS Data: Average Growth in Real Consumption ($K = 1$)

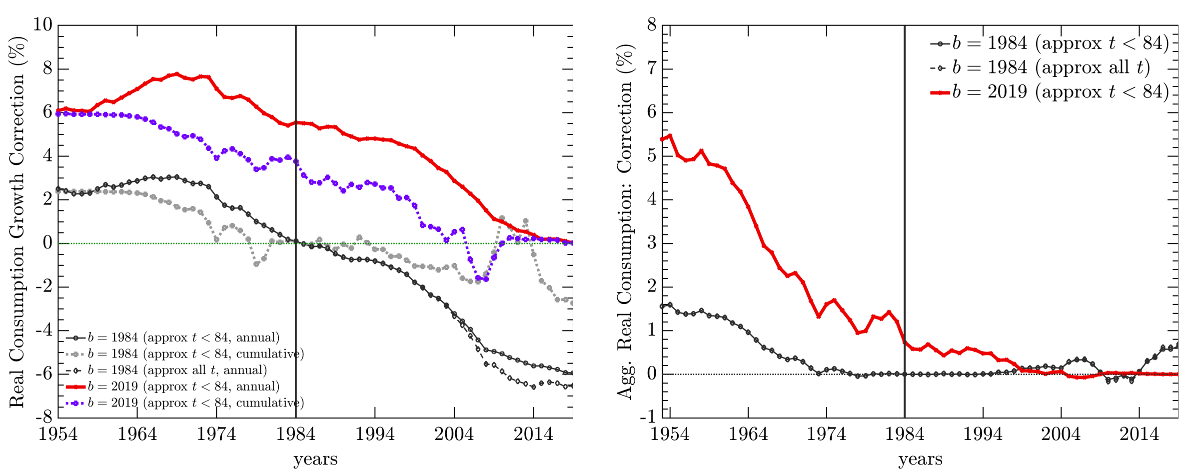
(a) Initial Period as Base ($b = 1984$)　　(b) Final Period as Base ($b = 2019$)




*Note:* The figure shows the evolution the average corrected and uncorrected measures of real consumption across quintiles relative to the measure of aggregate real consumption that ignores income heterogeneity. The latter defines the reduced-form index of real consumption using aggregate consumption expenditure shares. Panels (a) and (b) show the correction for the initial and final years as the base period, respectively, with $K = 1$. The reduced-form prices indices used for the 2nd and 1st order approximations are geometric and Tornqvist indices, respectively.

## Figure A7: CEX and BLS Data: Correction Over Longer Horizon ($K = 2$)

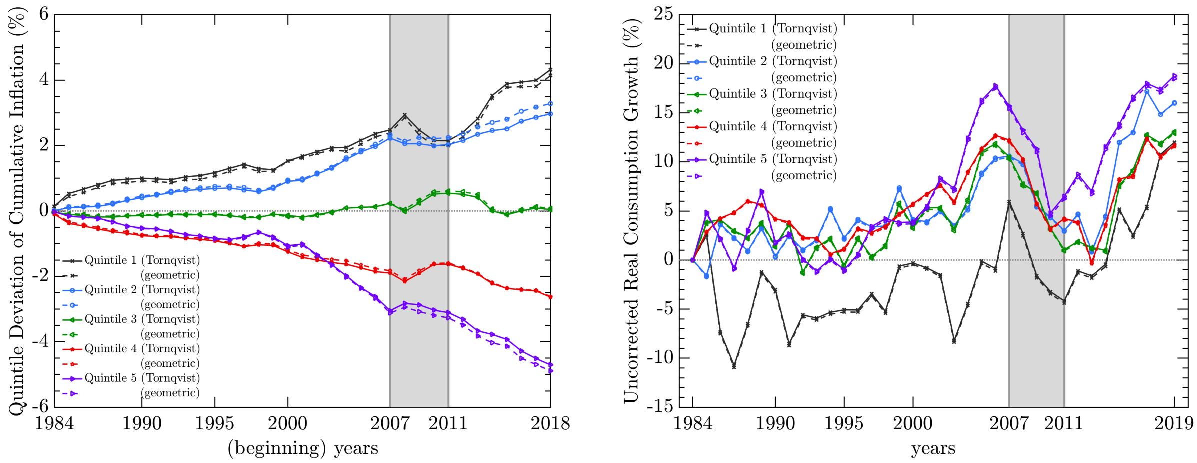
(a) Nonhomotheticity Growth Bias　　(b) Real Consumption Correction




*Note:* Panels (a) and (b) show the evolution of the nonhomotheticity correction and the corrected relative to uncorrected index of real consumption using aggregate consumption expenditure shares.

## Figure A8: BLS Data: Conventional Price and Quantity Indices

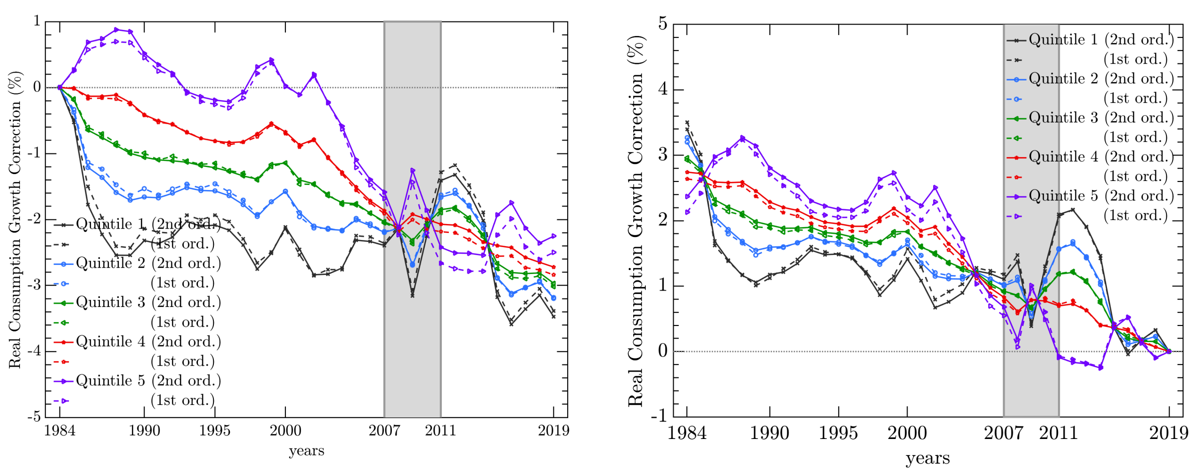
(a) Quintile Deviation of Price Indices (b) Quintile Quantity Indices




*Note:* Panels (a) and (b) show the evolution of cumulative reduced-form price and quantity indices for each quintile of income over the period, respectively. The great recession has been indicated in grey background.

## Figure A9: BLS Data: Bias in Reduced-form Real Consumption Growth ($K = 2$)

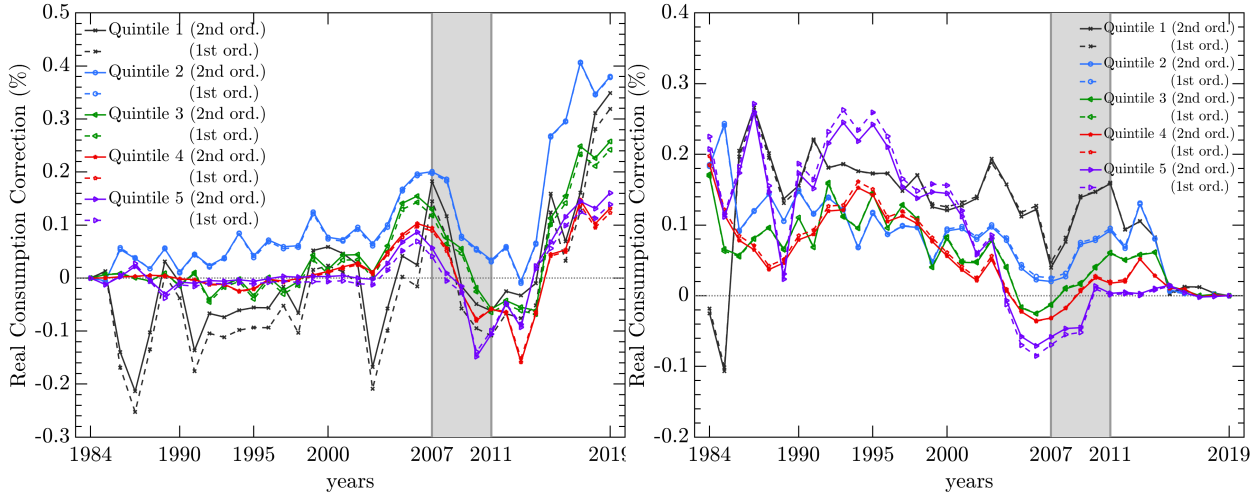
(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the evolution of the annual bias in the reduced-form measures of real consumption growth $\lambda_{b,t}^{n}$, defined in Equation (28), for different quintiles of income for the initial and final years as the base period.

Figure A10: BLS Data: Corrected relative to Uncorrected Real Consumption ($K = 2$)

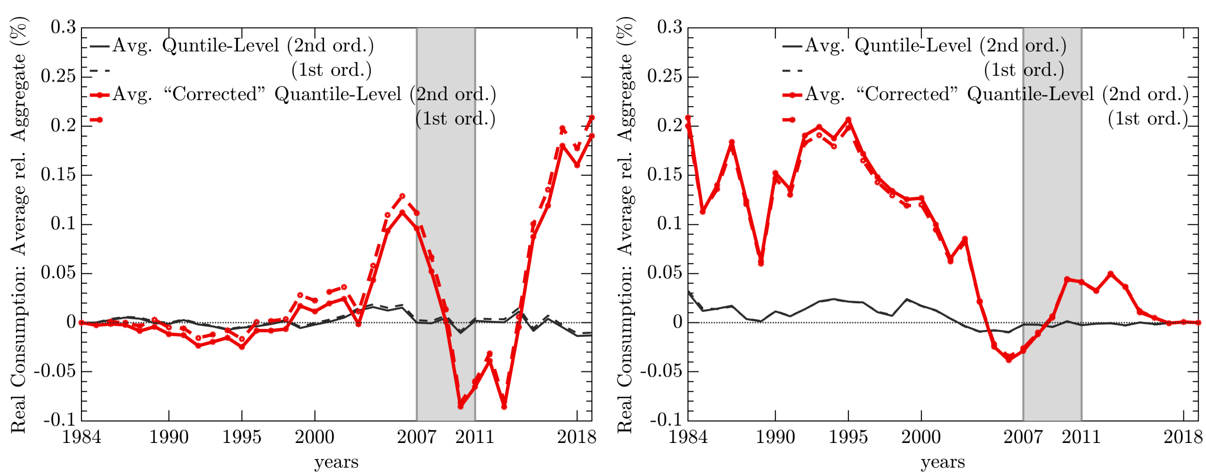
(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the ratio of the corrected to the uncorrected measures of real consumption vary over time for each quintile of initial real consumption for the initial and final years as the base period, respectively. The great recession has been indicated in grey background.

Figure A11: BLS Data: Average Growth in Real Consumption ($K = 2$)

(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)

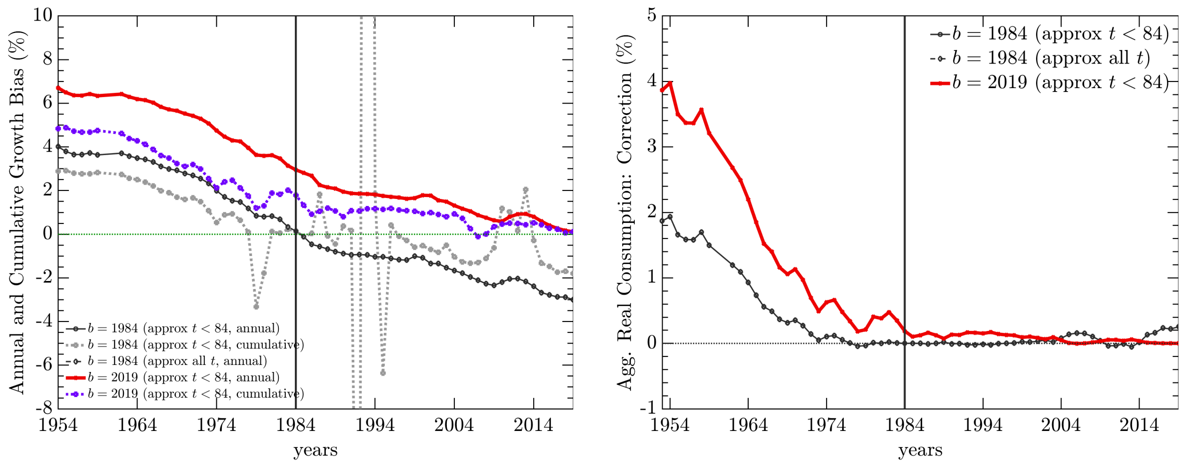



*Note:* The figure shows the evolution the average corrected and uncorrected measures of real consumption across quintiles relative to the measure of aggregate real consumption that ignores income heterogeneity. The latter defines the reduced-form index of real consumption using aggregate consumption expenditure shares. Panels (a) and (b) show the correction for the initial and final years as the base period, respectively. The reduced-form prices indices used for the 2nd and 1st order approximations are geometric and Tornqvist indices, respectively.

## Figure A12: BLS Data: Correction Over Longer Horizon ($K = 1$)

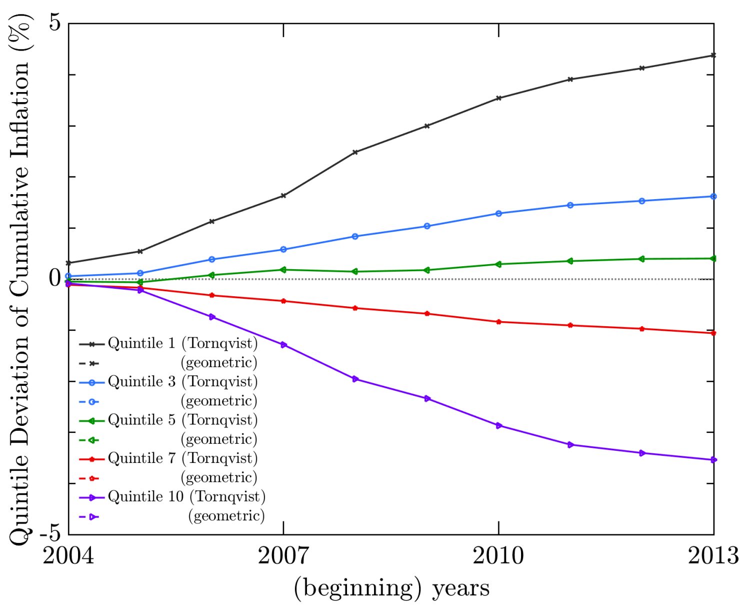
(a) Nonhomotheticity Growth Bias

(b) Real Consumption Correction




*Note:* Panels (a) and (b) show the evolution of the nonhomotheticity correction and the corrected relative to uncorrected index of real consumption using aggregate consumption expenditure shares.

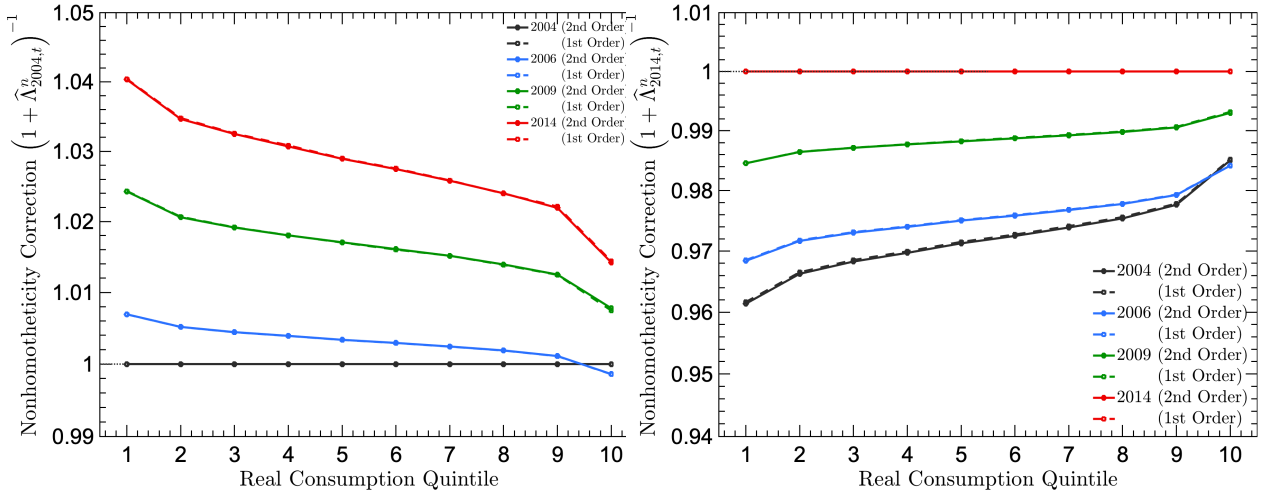
## Figure A13: Nielsen Data: Price Indices




*Note:* The figure shows the evolution of cumulative reduced-form price indices for each quintile of income over the period. The great recession has been indicated in grey background.

Figure A14: Nielsen Data: Nonhomotheticity Correction ($K = 2$)

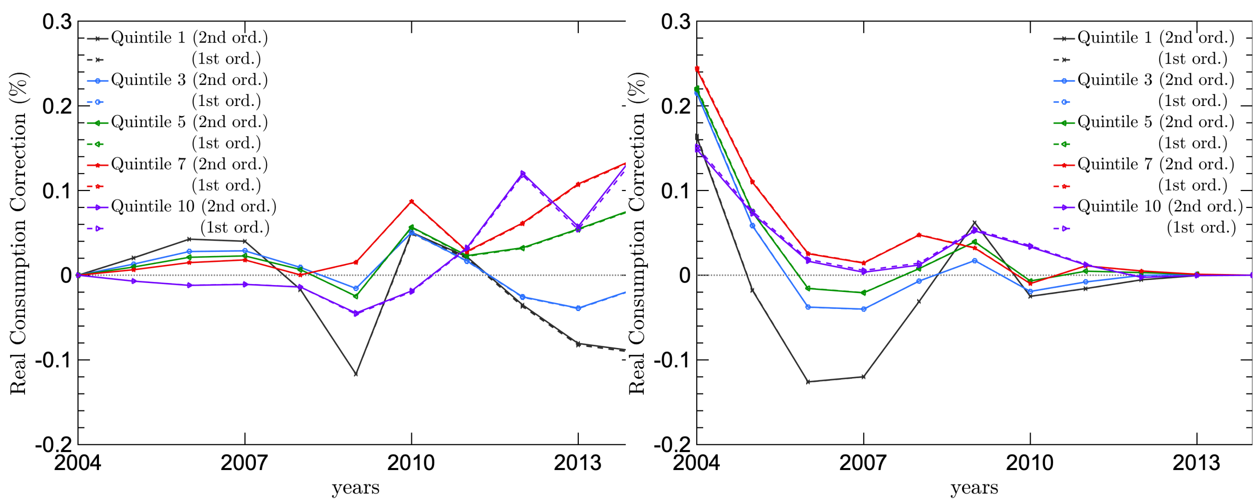
(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the "Nonhomotheticity Correction" varies over time for different quintiles of income for the initial and final years as the base period.

Figure A15: Nielsen Data: Corrected relative to Uncorrected Real Consumption ($K = 2$)

(a) Initial Period as Base ($b = 1984$) (b) Final Period as Base ($b = 2019$)




*Note:* Panels (a) and (b) show how the ratio of the corrected to the uncorrected measures of real consumption vary over time for each quintile of initial real consumption for the initial and final years as the base period, respectively. The great recession has been indicated in grey background.