the review of income and wealth

Review of Income and Wealth
Series 67, Number 3, September 2021
DOI: 10.1111/roiw.12495

MULTILATERAL METHODS, SUBSTITUTION BIAS, AND CHAIN DRIFT: SOME EMPIRICAL COMPARISONS

BY DANIEL MELSER*

Monash University

AND

MICHAEL WEBSTER

Australian Bureau of Statistics

Multilateral methods are increasingly being used in the computation of official price indexes such as the consumer price index (CPI). This reflects the growing use of scanner data by statistical agencies and the fact that fixed-based or chained price comparisons perform poorly in this context. A range of multilateral approaches have been pursued by different statistical agencies. Yet, it can be shown that some of these methods, at least theoretically, could suffer from substitution bias. We investigate this as well as the drivers of chain drift. We adopt a simulation-based approach using actual scanner data prices with the corresponding quantities being generated assuming constant elasticity of substitution (CES) preferences. We find that most methods systematically deviate from the exact CES benchmark index. This is even the case for the superlative index methods, which should not exhibit substitution bias. Interestingly, we also find significant chain drift even in the exact CES indexes. We argue this reflects life cycle pricing and particularly run-out sales at the end of a product's life.

JEL Codes: C43, C53, E31

Keywords: CES, consumer price index (CPI), multilateral indexes, scanner data, simulation

1. Introduction

In recent years, electronic point of sale data, or scanner data for short, has increasingly become available. This provides data on actual transactions—sale prices and quantities for individual products—collected by retailers at the point of sale. These data, of course, offer great potential to statistical agencies interested in measuring price changes for goods. The conventional approach has statistical agencies sending staff into stores to write down prices on clipboards for specific items at certain times of the sampling cycle. Scanner data offer the entire panoply of data and therefore presents an opportunity to greatly increase the quality and accuracy of official statistics. Moreover, it potentially provides this at lower cost

Note: Note that the views expressed do not represent those of the Australian Bureau of Statistics. The authors are grateful to Mike Krueger and IRI for making available the scanner data used in this study. The paper was improved as a result of a number of useful comments received from Jan de Haan, the editor and three anonymous referees. Any remaining errors are our own.

*Correspondence to: Daniel Melser, Monash University, 900 Dandenong Road, Caulfield East, VIC 3145, Australia (daniel.melser@monash.edu).

in terms of reducing the requirement to have field-staff visit retailers. This also reduces the burden on stores of having price collectors visit.

Although the apparent benefits are significant, the use of scanner data in official statistics, such as the consumer price index (CPI), has evolved relatively slowly. This has reflected the fact that new methods have needed to be developed to deal with this type of data. In particular, the fact that both price and quantity data are now available at a highly disaggregated level has posed challenges. First, the turnover or churn in products appearing and disappearing from the market is significant. This means that many prices are not matched from period-to-period. Second, very significant volatility in sales quantities is observed as a result of sporadic price discounting by retailers.[1]

To address these issues, Ivancic *et al*. (2011) proposed a novel multilateral method for constructing price indexes. They called this the Rolling Year GEKS (RYGEKS) method.[2] This approach calculated a transitive set of price indexes over a span of time—they suggested 13 months. This partly ameliorates the issue of changing product ranges, as the proportion of new and disappearing goods is likely to be smaller over a 13-month period than over, say, a handful of years. They argued that the key benefit of using multilateral methods was that the index would not suffer from "drift." This is a property that they contended primarily arose from stockpiling behavior by households during sales. The stockpiling argument goes as follows. When a product is discounted, purchases spike. In the following period, when prices return to their pre-discounted level, purchases are abnormally low. This is because some of the previously purchased items have not yet been consumed. This stockpiling behavior has the effect of giving greater weight to the price fall than the price rise. It has been shown that this can lead to significant downward drift in chained price indexes—where one period is compared with the next (de Haan and van der Grient, 2011).

Since the seminal contribution of Ivancic *et al*. (2011), many other approaches, some old and some new, have been proposed for the calculation of price indexes in this context. We consider two main alternative approaches to those proposed by Ivancic *et al*. (2011). The first approach we consider is what we call the temporal Geary–Khamis (GK) method (Geary, 1958; Khamis, 1972; Chessa, 2016). This constructs a set of reference prices for each product and then calculates the price level by comparing the prices in each period to these reference prices. The second approach we consider is the time-product dummy (TPD) method. This approach is based on the specification of a regression model for prices—dummy effects are included for each time and product. The price index is calculated from the estimated coefficients on the time-dummy variables (Rao and Baneerjee, 1986; Aizcorbe et al., 2003; Krsinich, 2016). Rao (2005) showed that this is also a reference price method. The main difference from the GK method is the use of geometric, as opposed to arithmetic, averages in constructing the references prices and in the index number formula.

[1]An additional issue is that the sales quantities reflect purchases by consumers as well as nonresidents, government, and business. As noted by a referee, from a CPI perspective the nonconsumer purchases should be removed.

[2]This built on earlier work looking at the use of the GEKS approach in the context of temporal indexes such as Balk (1981) and Kokoski *et al*. (1999).

The present state of the literature reflects a degree of uncertainty about which of these methods is to be preferred. On one hand, the approaches appear to yield quite similar results empirically and are fairly robust to the various choices required to implement them (ABS, 2016; Melser, 2018). Yet, on the other hand, these various approaches are conceptually quite different and therefore appear to have the potential to yield quite different results. One area where the contrast between the methods is perhaps the strongest is in the underlying economic basis of the approaches (Diewert and Fox, 2018). The approach of Ivancic *et al*. (2011) uses superlative index numbers, such as the Fisher and Törnqvist indexes, and aggregates these flexibly using the GEKS method in the spirit of Caves *et al*. (1982).[3] This approach is unlikely to suffer from substitution bias. However, whether the other approaches will suffer from substitution bias is less clear.

This paper has two main objectives. First, to try and shed some light on the extent of substitution bias in various multilateral price index methods. In this regard, we take Diewert and Fox (2018) as our starting point and make use of their framework. They provide some theoretical results that show that some of the proposed methods may be susceptible to substitution bias. They also present some results from a small simulation model. This is based on using certain assumptions about consumer behavior to generate the quantities demanded, based on the prices they hypothesize. They use this simulated data to calculate a range of multilateral and bilateral indexes. They illustrate that certain multilateral methods can yield results that diverge from the true index in some circumstances. Our primary focus is to build on the simulation framework developed by Diewert and Fox (2018) and develop a more comprehensive simulation architecture to derive more precise results about substitution bias.

Second, we use our simulation framework to explore the drivers of chain drift. As we noted previously, the main explanation for chain drift has been stockpiling by consumers. However, we identify chain drift even in our simulation framework—this is based on a classical demand model and therefore does not allow for stockpiling by construction. We find that life cycle pricing, in particular run-out sales, is an important driver of chain drift.

One of the challenges in constructing such a simulation is to balance realism against our ability to interpret the results. We follow Diewert and Fox (2018) in supposing that consumers have constant elasticity of substitution (CES) preferences over products. This is a functional form that is particularly attractive in this context as it provides a clear means of manipulating the degree of substitution. It is also attractive because the exact index number formulae in the case of CES preferences are well known. This means we know what true price change is for our simulated data. This can be used as a yardstick to compare the performance of other methods.[4]

[3]See Diewert (1976) for the precise definition of superlative index numbers. In short, superlative indexes are derived from a functional form that provides a second-order approximation to consumer preferences. Therefore, such indexes account very well for patterns of substitution between products.

[4]We also note that there are more flexible representations of preferences than those embodying the CES property (see Diewert, 1976). For example, de Haan (2019) argued that results from assuming the CES functional form should be treated with caution.

One innovation we adopt is to use actual price data from the IRI data set (Bronnenberg *et al.*, 2008). This is an important point of difference with Diewert and Fox (2018), who construct their own price data. An alternative would be to construct a stochastic model of prices. Our approach of using actual data is desirable for several reasons. First, it is clear that the price data are realistic—it is actual data after all. Second, constructing a stochastic model of prices is likely to be quite challenging. Products go on sale often; prices drop significantly when they go on sale; a given product's price interacts in complex ways with other prices; and products also disappear from the market after a time. Accounting for these factors in a stochastic model would be challenging.

The use of actual price data means our approach is not a pure simulation—in the sense that the data are wholly synthetic—but it is also not truly empirical, as we assume a demand generation process. Our approach sits somewhere between these two extremes. We use actual price data to create two types of data sets. The first is a panel of prices. This uses actual prices and splices on prices as products disappear from the market. This creates a data set of prices without any missingness. In the second data set, we take a random sample of products from the primary data. We create these two data sets for each combination of the cities: Dallas, Los Angeles, and New York; and the products: beer, carbonated beverages, coffee, margarine and butter, peanut butter, soup, toilet tissue, and toothpaste.

Our simulation results provide several insights. First, we find a surprising level of divergence between the exact CES index and the Fisher and Törnqvist indexes. This is surprising because a priori we would expect the Fisher and Törnqvist indexes to approximate the exact CES indexes fairly closely. However, on average we find that the Fisher and Törnqvist GEKS indexes underestimate price change. Second, our results show a degree of upward bias in the TPD and GK methods in both the panel and sample data sets. This is expected based on the theory (Diewert and Fox, 2018). The bias is nontrivial. In the panel data set, it peaks at around 2 percentage points per annum when the elasticity of substitution in the CES functional form is 5 and decreases at higher elasticities. However, in the sample data set the bias gets larger as the elasticity rises. We calculate the bias up to an elasticity of 15. Here the bias reaches an alarming 5 percentage points per year. Third, we find in the sample-of-prices data set, which includes missingness, that there is significant chain drift even in the exact CES index. This is surprising given the prevailing view that it is stockpiling which generates chain drift. It is obviously not stockpiling that is generating these results as we construct demand using the CES framework that assumes simultaneous purchase and consumption. Given this, we argue that an important driver of chain drift is life cycle pricing and especially run-out sales of products at the end of their life. We show algebraically how this can lead to downwards drift. We illustrate empirically how, when we drop the last price observation for each product and therefore eliminate these run-out sales to an extent, that chain drift diminishes. Interestingly, this also significantly reduces the bias observed in the TPD and GK methods. This indicates that these methods may be somewhat susceptible to bias in the presence of life cycle pricing. These results are useful because the influence of life cycle pricing on chain drift and multilateral methods has not been well appreciated in the literature.[5]

In the next section, we outline our methodology and detail the various multilateral index number methods that we examine. In Section 3, we outline our data. Section 4 discusses our results in detail. In Section 5, we further explore the effects of life cycle pricing on index bias. Finally, in Section 6, we provide some brief conclusions.

2. Methodology

2.1. *Simulation Framework*

Our dual objectives are to understand the extent to which different methods may exhibit substitution bias and examine how they perform in the presence of new and disappearing goods. Our methodology to investigate this issue is simple, but we believe compelling. We use actual price data and assumptions about preferences, in particular the degree of substitution between products, to construct the quantities purchased, and expenditure shares. These data on prices and quantities/expenditures are then used to calculate a range of multilateral indexes that are compared with the known exact CES index. Let us first turn to consumer preferences.

We suppose that consumers have CES preferences.[6] That is, their preferences can be described by the following cost function:

$$C(p_t, U) = \left(\sum_{i \in I_t} a_i p_{it}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} U. \tag{1}$$

Here prices for product $i \in I_t$ available in period t are denoted p_{it}, I_t is the index set of available products, a_i is the measure of product quality (which we will set to 1 in our simulations), and U is the level of utility. The parameter σ is particularly important. It governs the degree of substitution between products. As shown below, it reflects the degree to which relative purchases move between goods as their relative prices change. Clearly σ must be nonnegative.

$$\frac{\partial \ln \left(\frac{x_{it}}{x_{jt}} \right)}{\partial \ln \left(\frac{p_{it}}{p_{jt}} \right)} = -\sigma. \tag{2}$$

[5]Although we note that the influence of life cycle pricing on index numbers has been better understood and appreciated by practitioners working at statistical agencies. In many cases they have taken steps to limit bias resulting from life cycle pricing trends in official indexes. See, for example, paragraph 2.19 of ABS (2017).

[6]CES preferences were introduced into the economics literature by Arrow *et al*. (1961).

Equation (2) is derived by making use of Shepard's Lemma, $x_{it} = \frac{\partial C}{\partial p_{it}}$, where x_{it} is the quantity of good i consumed in period t. In the case of CES preferences, Shepard's Lemma has the form

$$\frac{p_{it} x_{it}}{\sum_{i \in I_t} p_{it} x_{it}} \equiv s_{it} = \frac{a_i p_{it}^{1-\sigma}}{\sum_{i \in I_t} a_i p_{it}^{1-\sigma}}. \tag{3}$$

This is the equation that we use to generate demand in our simulations. Given prices, σ and also quality parameters (a_i) for each product, we can construct demand. Moreover, by manipulating σ we can generate demand functions that embody more or less aggressive substitution across products. Higher values of σ lead to greater substitution between products as relative prices change and vice versa.

The CES functional form is particularly useful because there is such a clear link between the degree of substitution and its underlying parameters. It is also useful because there are several well-known exact price indexes for CES preferences. These indexes are of course all equal if the consumer has CES preferences. We will note two exact CES indexes. The first is the Sato–Vartia index (Sato, 1976; Vartia, 1976). This takes the form

$$P_{vt}^{SV} = \prod_{i \in I_{vt}} \left(\frac{p_{it}}{p_{vi}} \right)^{w_{ivt}}, \; w_{ivt} = \frac{\left(\frac{\tilde{s}_{it} - \tilde{s}_{iv}}{\ln \tilde{s}_{it} - \ln \tilde{s}_{iv}} \right)}{\sum_{i \in I_{vt}} \left(\frac{\tilde{s}_{it} - \tilde{s}_{iv}}{\ln \tilde{s}_{it} - \ln \tilde{s}_{iv}} \right)}, \; \tilde{s}_{it} = \frac{p_{it} x_{it}}{\sum_{i \in I_{vt}} p_{it} x_{it}}. \tag{4}$$

The other index is a version of the quadratic-mean-of-order-r index discussed in Diewert (1976) and is shown below:

$$P_{vt}^{\sigma} = \left[\left(\sum_{i \in I_{vt}} \tilde{s}_{iv} \left(\frac{p_{it}}{p_{iv}} \right)^{(1-\sigma)} \right)^{\frac{1}{(1-\sigma)}} \left(\sum_{i \in I_{vt}} \tilde{s}_{it} \left(\frac{p_{it}}{p_{iv}} \right)^{-(1-\sigma)} \right)^{-\frac{1}{(1-\sigma)}} \right]^{\frac{1}{2}}. \tag{5}$$

Here $I_{vt} = I_t \cap I_v$ is the set of matched products across time periods v and t. Note that while (5) depends on σ interestingly, the Sato–Vartia index, equation (4), does not. When the demands are generated from CES preferences, these two indexes will provide the same answer.

As noted in the introduction, in some of our simulations we use actual price data where there are changes in the assortment of products available over time, that is, $I_t \neq I_v$. With new and disappearing goods, it is not quite correct to regard (4) or (5) as exact cost of living indexes. As Feenstra (1994) noted, when the product range changes over time, the cost of living will be influenced by these developments. That is, the cost-of-living index can be decomposed as

$$\frac{C(p_t, U|I_t)}{C(p_v, U|I_v)} = \left[\frac{C(p_t, U|I_t)}{C(p_t, U|I_{vt})} \frac{C(p_v, U|I_{vt})}{C(p_v, U|I_v)} \right] \frac{C(p_t, U|I_{vt})}{C(p_v, U|I_{vt})}. \tag{6}$$

The term in square brackets on the right-hand side of (6) is an effect related to changes in variety. In this work our focus is on measuring matched price change, and we do not attempt to account for the effects of new and disappearing goods.[7] The final term on the right-hand side is a matched price index based on products sold in both periods. The indexes in (4) and (5) are exact for this matched products price index and it is this index that is our focus.

In the case of new and disappearing goods, the CES exact bilateral matched model indexes (4) and (5) will not be transitive, for example, $P_{vt}^{SV} \neq P_{rt}^{SV} \times P_{vr}^{SV}$. We resolve this by using the weighted GEKS (WGEKS) approach to aggregate each of the bilateral indexes into a transitive multilateral index (see, for example, Rao and Timmer, 2003). The GEKS approach estimates a transitive set of indexes by minimizing the least-squares distance between the bilateral indexes and a set of transitive indexes (Gini, 1931; Eltetö and Köves, 1964; Szulc, 1964). For our particular application of this method, we will extend this approach and use, what we dub, a "double-GEKS" (DGEKS) approach to aggregation.

In the seminal work of Ivancic et al. (2011), a key feature of the method they proposed was the use of a rolling window over which parities—that is, the relative price levels between time periods—were calculated. The parities for each new window were then linked to the existing indexes calculated from previous windows. The use of a rolling window ensured that there was a reasonable degree of product matching in the periods compared and that the weights represented contemporary consumption patterns. In our simulations, we try to preserve as much of the spirit of this as possible. However, some care must be taken here. The linking methods are aimed at forcing the parities calculated over the new window to align with the old parities so the index does not need to be revised. Our focus on the simulations is a little different. We are trying to detect substitution bias, and imposing the non-revisability constraint might make this more difficult. Instead, our approach is the following.

First, in the spirit of Ivancic *et al.* (2011), we calculate parities over each of the 13-month windows using a given bilateral index and the WGEKS approach. With regard to equation (7), we estimate the β's over some set of ρ time periods T_u^{ρ}—here subscript u denotes the first period of the rolling window and ρ the length of the window. In our simulations we set ρ = 13 months. This is done for all rolling windows; $T_1^{\rho}, T_2^{\rho}, \ldots, T_{T+1-\rho}^{\rho}$. This yields a set of parities β^u for each of the windows.

$$\text{(7)} \qquad \text{SSE}(\beta^u) = t T_u^{\rho} \in \tilde{\sum} v T_u^{\rho} \in \tilde{\sum} w_{vt} (\ln P_{vt} - (\beta_t^u - \beta_v^u))^2.$$

In our simulations below, P_{vt} is the Sato–Vartia index and w_{vt} are the weights applied in the WGEKS aggregation. The weights we will use are the average matched expenditure share (AMES) weights from Melser (2018). These weights reflect the average proportion of expenditure that is made up from products that are available in both periods. In the case where there are no missing products, then

[7]A literature has developed since Feenstra (1994) that estimates the size of the variety effect using the decomposition in equation (6). See, for example, Melser (2006), Broda and Weinstein (2010), and Melser (2019). A requirement of this approach is that the elasticity of substitution is greater than one.

the weights are all one and WGEKS will be equivalent to GEKS. More generally the AMES weights are

$$w_{vt} = \sum_{i \in I_{vt}} \left(\frac{s_{it} + s_{iv}}{2} \right). \tag{8}$$

To calculate an overall index, we need to aggregate these window parities into a single series. We do this by running GEKS again. First, we reconstruct each of the bilateral comparisons from the window parities. For example, for window T_u^ρ we calculate $\ln P_{tv}^u = \beta_t^u - \beta_v^u$ for each t and v in T_u^ρ. Second, these indexes are then used in a second round of GEKS aggregation.

$$\mathrm{SSE}(\tau) = \sum_{u=1}^{T+1-\rho} \left(\sum_{t \in T_u^\rho} \sum_{v \in T_u^\rho} (\ln P_{tv}^u - (\tau_t - \tau_v))^2 \right). \tag{9}$$

This DGEKS approach yields a transitive set of parities, τ. We also note that the second step of DGEKS, shown in equation (9), can be used with any particular inputs for $\ln P_{tv}^u$. That is, they need not come from a first-stage WGEKS approach but could come from another multilateral method. Indeed, we will also use this approach to aggregate the parities for each window for the TPD and GK methods.

2.2. *The Multilateral Methods Considered*

We will compare the following three main alternative approaches against the indexes outlined in the previous section: (1) WGEKS using alternative bilateral index number formulae, (2) the TPD method, and (3) the temporal GK approach.

The GEKS approach to aggregating bilateral price indexes, reflected in (7), can be applied with any particular bilateral index number formula. In particular, we will consider two well-known bilateral index number formulae: the Fisher and Törnqvist indexes.[8] These are shown respectively below:

$$P_{vt}^F = \left(\sum_{i \in I_{vt}} \tilde{s}_{iv} \left(\frac{p_{it}}{p_{iv}} \right) \right)^{\frac{1}{2}}, \left(\sum_{i \in I_{vt}} \tilde{s}_{it} \left(\frac{p_{it}}{p_{iv}} \right)^{-1} \right)^{-\frac{1}{2}} \tag{10}$$

$$P_{vt}^T = \prod_{i \in I_{vt}} \left(\frac{p_{it}}{p_{iv}} \right)^{\frac{\tilde{s}_{it} + \tilde{s}_{vt}}{2}}. \tag{11}$$

This approach is conceptually quite similar to our exact approach. The prevailing wisdom is that this approach should approximate the exact CES approach

[8]The use of the Törnqvist index and the GEKS approach is known as the CCD method and originates from Caves *et al*. (1982).

quite closely. This is because both the Fisher and Törnqvist indexes are superlative indexes in the sense that they can be derived from a flexible (second-order) representation of consumer preferences (Diewert, 1976).

We can also use these bilateral formulae to calculate direct indexes between the first period and all other periods. This provides a transitive set of price indexes. However, this approach is only sensible in the case of the panel-of-prices data set. Here there are no missing prices. In the other case we consider, the sample-of-prices data set, direct indexes are likely to be negatively affected by lower product matching rates as time progresses.

The second approach, the TPD method, is more substantively different. It dates back to Summers (1973) who suggested using a regression model to fill "holes" (i.e., missing prices) in cross-country price data. This model included fixed effects for country and for product and therefore was dubbed the country-product-dummy method or CPD. TPD represents the time series adaptation of this approach (Aizcorbe et al., 2003; Krsinich, 2016). The TPD method is a regression model with fixed effects for each product and time.

$$\ln p_{it} = \alpha_i + \delta_t + \varepsilon_{it}. \tag{12}$$

In this approach, we estimate δ_t by least squares. If each equation is weighted, using weights w_{it}, then this gives the weighted TPD (WTPD) method. The unweighted TPD method is not widely used in practice because it is intuitively clear that some weighting should be used to account for the differential importance of products. Therefore, we focus on WTPD. The most common weighting strategy is to use expenditure shares, s_{it} (Diewert, 2005). This will be the approach that we follow. The index between any two periods is then defined as $P_{vt}^{WTPD} = \exp\left(\delta_t - \delta_v\right)$.

In implementing the WTPD approach, we use a 13-month rolling window and the second stage of the DGEKS procedure outlined earlier to derive the overall parities. That is, in the first stage we estimate equation (12) over 13-month rolling windows to obtain many WTPD parities. In the second stage, we apply the GEKS method as in (9) to derive the final parities.

The final approach we consider is the temporal version of the well-known GK method (Geary, 1958; Khamis, 1972). The GK method has historically been widely used in constructing international comparisons of prices and income such as in the Penn World Tables (Feenstra *et al.*, 2015). It has been adapted to the temporal context recently by Chessa (2016). It calculates the price index relative to a reference set of prices for each product. Denote the reference prices, $\bar{p}_i$. Two periods, t and v, are compared via the reference prices using the quantities sold in each period. This amounts to the ratio of two Paasche-like indexes.

$$P_{vt}^{GK} = \frac{\left(\frac{\sum_{i \in I_t} p_{it} x_{it}}{\sum_{i \in I_t} \bar{p}_i x_{it}}\right)}{\left(\frac{\sum_{i \in I_v} p_{iv} x_{iv}}{\sum_{i \in I_v} \bar{p}_i x_{iv}}\right)}. \tag{13}$$

This approach actually defines a family of indexes called Generalized Unit Value indexes (de Haan, 2007; von Auer, 2014). These differ based on how the various $\overline{p}_i$ are chosen. However, we will just consider the GK approach. In this case we have

$$\overline{p}^{\rho}_{ui} = \sum_{t \in T^{\rho}_{ui}} \left(\frac{x_{it}}{\sum_{t \in T^{\rho}_{ui}} x_{it}} \right) \frac{p_{it}}{P^{GK}_{ut}}. \tag{14}$$

Here T^{ρ}_{ui} is the index set of periods in which product i is sold in the rolling window of periods T^{ρ}_{u}. It can be observed that the reference price is the quantity-weighted "real" price—that is, the price adjusted for the price level in period t relative to some period such as the first period of the window, u.

In implementing the GK approach, we use a 13-month rolling window and the second stage of the DGEKS to derive the overall parities. This is the same as for the WTPD method. First, we estimate parities for each of the 13-month rolling windows. Second, we apply GEKS in the second stage as in (9) to derive the final parities.

2.3. *Expectations from Theory*

In this section, we consider the theoretical expectations regarding substitution bias for the methods outlined earlier. This draws upon Diewert and Fox (2018) as well as others.

2.3.1. Fisher and Törnqvist WGEKS

The Fisher and Törnqvist WGEKS approaches are expected to provide estimates of price change, which are free of substitution bias. This reflects the fact that the Fisher and Törnqvist bilateral indexes can be justified on the basis of flexible functional forms that allow for complex patterns of substitution between goods. In fact, these indexes are based on more flexible representations of consumer preferences than the CES functional form. Diewert (1976) showed that the underlying preferences can provide a second-order approximation to the consumer's cost function at a given point. The CES functional form is not this flexible.

However, this does not necessarily mean that the multilateral indexes produced by the Fisher and Törnqvist WGEKS and direct approaches will be similar to those using CES preferences. It is worth noting some of the similarities and differences between the CES, Fisher, and Törnqvist bilateral indexes. The CES index shown in (5) is actually part of a wider family of indexes called quadratic-mean-of-order-r indexes. These can be obtained by substituting $\sigma = 1 - \frac{r}{2}$ into (5). This gives

$$P^{r}_{vt} = \left(\sum_{i \in I_{vt}} \tilde{s}_{iv} \left(\frac{p_{it}}{p_{iv}} \right)^{\frac{r}{2}} \right)^{\frac{1}{r}} \left(\sum_{i \in I_{vt}} \tilde{s}_{it} \left(\frac{p_{it}}{p_{iv}} \right)^{-\frac{r}{2}} \right)^{-\frac{1}{r}}. \tag{15}$$

It can be seen that when $r = 2$ ($\sigma = 0$) we obtain the Fisher index, and when $r = 0$ ($\sigma = 1$) it can be shown that we obtain the Törnqvist index, that is, the Fisher index will equal the CES index in the case of $\sigma = 0$ and Törnqvist will equal the CES index when $\sigma = 1$. However, for other values of σ (or r), it is not as clear what the relativities of the indexes will be.

2.3.2. The Geary–Khamis Method

Diewert and Fox (2018) discussed the GK method and its economic justification at length. They note that the GK index will be exact in two cases. Indeed, Diewert (1999) showed that these were the only cases in which GK was exact.

First, when consumers have a linear utility function, $U_t = \sum_{i \in I_t} \frac{x_{it}}{b_i}$. In the linear utility function, each of the products is a perfect substitute for each other after adjusting by $b_i > 0\ \forall i$. The cost function dual to this utility function takes the form $C(p_t, U_t) = min\left\{b_i p_{it}, i \in I_t\right\}$. This cost function will equal the CES cost function when $a_i = b_i$ and σ is infinite.

Second, when consumers exhibit no substitution between products. That is when $U_t = min\left\{\frac{x_{it}}{c_i}, i \in I_t\right\}$ with $c_i > 0 \forall i$. In this case consumers purchase products in fixed proportions. That is, there is no substitution between products. The cost function dual to this utility function has the form $C(p_t, U_t) = \sum_{i \in I_t} c_i p_{it}$. This will equal the CES cost function when $a_i = c_i$ and $\sigma = 0$.

Thus the GK method will be exact in two extreme cases—when $\sigma = 0$ and when σ is infinite. In the intervening range of σ values, it appears likely that the GK method will be upwardly biased, as it does not allow for moderate levels of substitution. However, this leaves open the extent of this bias and the point at which the bias is largest.

2.3.3. Weighted TPD

To consider the issue of substitution bias in the TPD method, note that if we take logs of Shepard's Lemma, equation (3), and rearrange, we get the following:

$$\ln p_{it} = \frac{1}{(1-\sigma)} \ln s_{it} - \frac{1}{(1-\sigma)} \ln a_i + \ln\left(\sum_{i \in I_t} a_i p_{it}^{1-\sigma}\right)^{\frac{1}{(1-\sigma)}}. \tag{16}$$

This can be easily converted into an estimable model of the form below:

$$\ln p_{it} = \frac{1}{(1-\sigma)} \ln s_{it} + \alpha_i^* + \delta_t^*, \tag{17}$$

where $\alpha_i^* = \frac{1}{(1-\sigma)} \ln a_i$ is the product-dummy effect and δ_t^* is the time-dummy effect reflecting the log of the unit cost function $\ln\left(\sum_{i \in I_t} a_i p_{it}^{1-\sigma}\right)^{\frac{1}{(1-\sigma)}}$. Clearly this model is intimately related to the TPD model in equation (12). The difference being that the term, $\frac{1}{(1-\sigma)} \ln s_{it}$, appears in (17) but is excluded from the TPD model.

In the specific case that $\sigma = 1$—when the cost function has the Cobb–Douglas form—then the TPD approach produces an exact index. This is because the expenditure shares are fixed across time, $s_{it} = s_i$, and the share in equation (17) gets absorbed into the product-specific intercept, α_i^*.

Diewert and Fox (2018) also noted that the TPD method is exact for a linear cost function. To see this suppose that $\sigma \rightarrow \infty$, then the CES cost function takes the form

$$C(p_t, U) = min\left\{a_i p_{it}, i \in I_t\right\}. \quad (18)$$

That is, there is perfect substitutability between the products, and therefore the consumer will choose the lowest-priced product (on a quality-adjusted basis) and expenditure on all other products is zero. This pattern of expenditure shares will arise out of (3) when σ becomes very large. In this case, the cost of living index can be represented simply by the weighted movement in prices such as in equation (12).

This means that the bias in the WTPD method will be zero when $\sigma = 1$ and also when σ is very large. Bias will likely tend to be positive in the more general case. This leaves open the size of the bias and the particular value of σ at which it peaks.

3. Data

As previously noted, we use actual price data in our simulation. This comes from the widely used IRI scanner data set introduced by Bronnenberg *et al.* (2008).[9] This is a large US scanner data set put together by IRI from the stores that provided it with data. The data set stretches from 2001 to 2012—a span of 12 years. The data set includes information on several cities and products. We focus on data for three large cities: Dallas, Los Angeles, and New York. In terms of products, we use data for eight categories: beer, carbonated beverages, coffee, margarine and butter, peanut butter, soup, toilet tissue, and toothpaste. These products have varying levels of stockpilability and purchase frequency and have differing importance in the household budget according to Bronnenberg *et al.* (2008; Table 2). In total, this gives 24 different city-product combinations for which we will compare simulation results. The data we use will be aggregated to a monthly frequency (from weekly) as this is the most common publication frequency for statistical agencies globally.

Table 1 summarizes the data from which our simulation data sets are derived. The data for each of the product-city combinations are significantly-sized and when taken together represent a massive volume of transactions. The total value of sales observed in the data is around \$4.3 billion over the 12 years. Across all products and cities, more than 39 million observations are included (monthly unit value price and quantity by barcode in a given store). This reflects data from 50,949 different products and 1,441,629 different store-product combinations. In our

[9]A full listing of the papers that have used these data as of December 2017 is provided by Kruger (2017).

TABLE 1
Scanner Data Summary Statistics (Full Data Set)

Product	City	No. Obs.	Tot. Exp. ($mil.)	No. Prods.	No. Stores	No. Prods × Stores	Price ($)	
							Mean	Std. Dev.
Beer	Dallas	954,765	216	2,166	93	33,549	8.96	4.78
	Los Angeles	2,864,393	590	2,283	147	85,624	9.32	5.54
	New York	1,317,470	193	3,867	180	49,525	8.65	4.39
Carbonated beverages	Dallas	2,122,219	303	3,980	101	84,899	2.22	1.43
	Los Angeles	3,722,751	600	3,026	147	133,281	2.40	1.51
	New York	4,256,475	618	5,390	186	156,649	2.15	1.42
Coffee	Dallas	1,027,513	49	3,131	101	45,529	5.36	2.68
	Los Angeles	2,205,683	135	2,742	147	90,966	6.38	2.75
	New York	2,215,967	216	4,266	186	88,811	5.86	3.04
Margarine and butter	Dallas	338,956	24	382	102	10,537	1.98	1.10
	Los Angeles	631,829	63	294	147	16,698	2.48	1.35
	New York	793,315	77	435	186	22,956	2.60	1.26
Peanut butter	Dallas	275,407	16	508	101	9,974	3.11	1.57
	Los Angeles	568,270	47	470	147	18,169	3.86	1.80
	New York	587,254	58	739	186	19,903	3.51	1.74
Soup	Dallas	1,697,864	59	3,017	99	61,950	1.79	0.87
	Los Angeles	3,579,498	159	2,819	144	114,883	2.07	1.02
	New York	4,342,255	214	3,915	184	147,945	2.08	0.93
Toilet tissue	Dallas	289,423	72	804	102	15,627	5.66	3.40
	Los Angeles	517,458	166	702	147	26,434	6.37	3.71
	New York	681,552	243	961	186	35,421	6.53	4.03
Toothpaste	Dallas	842,535	19	1,667	101	36,775	3.45	3.37
	Los Angeles	1,496,775	49	1,437	147	59,474	3.84	3.91
	New York	1,814,396	79	1,948	186	76,050	3.60	3.63
Total		39,144,023	4,264	50,949	3,453	1,441,629	—	—

TABLE 2
SIMULATION RESULTS

Method	Mean					Median					% With Positive Bias				
σ:	0	2	5	10	15	0	2	5	10	15	0	2	5	10	15
Panel of prices															
WGEKS															
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fisher	0.00	−0.42	−2.70	−2.43	−2.16	0.00	−0.21	−2.18	−1.96	−1.96	0.00	12.50	4.17	20.83	20.83
Törnqvist	0.03	−0.06	−1.57	−1.89	−1.79	0.02	−0.04	−1.21	−1.76	−1.76	87.50	4.17	0.00	20.83	20.83
Direct															
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fisher	0.00	−0.59	0.85	2.37	2.58	0.00	−0.45	−0.01	2.74	2.74	0.00	12.50	45.83	87.50	79.17
Törnqvist	0.08	−0.11	0.51	1.97	2.28	0.02	−0.07	0.07	2.11	2.11	79.17	12.50	54.17	79.17	75.00
Chained															
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fisher	0.00	−0.30	−1.48	−1.25	−1.14	0.00	−0.14	−1.12	−1.43	−1.43	0.00	16.67	29.17	29.17	33.33
Törnqvist	0.00	−0.05	−0.83	−0.91	−0.87	0.00	−0.02	−0.52	−0.95	−0.95	54.17	16.67	25.00	29.17	33.33
Geary–Khamis	0.00	0.20	1.90	1.29	0.80	0.00	0.10	1.57	0.90	0.90	0.00	87.50	100.00	87.50	70.83
WTPD	−0.04	0.09	1.43	1.02	0.65	−0.03	0.05	1.29	0.66	0.66	4.17	95.83	95.83	87.50	70.83
Sample of prices															
WGEKS															
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fisher	0.00	−0.15	−1.23	−1.67	−1.22	0.00	−0.15	−1.11	−1.18	−1.18	0.00	8.33	8.33	8.33	29.17
Törnqvist	−0.01	−0.02	−0.66	−1.39	−1.15	−0.01	−0.02	−0.59	−1.24	−1.24	8.33	8.33	8.33	8.33	25.00
Chained															
CES (Sato–Vartia)	−1.06	−1.77	−4.04	−5.82	−5.74	−1.03	−1.61	−3.61	−4.92	−4.92	20.83	16.67	8.33	20.83	25.00
Fisher	−1.06	−2.21	−6.96	−9.38	−8.98	−1.03	−1.98	−4.29	−6.44	−6.44	20.83	12.50	4.17	12.50	20.83
Törnqvist	−1.10	−1.84	−5.66	−8.58	−8.51	−1.08	−1.71	−4.10	−6.07	−6.07	20.83	16.67	4.17	12.50	25.00
Geary–Khamis	−0.19	−0.36	0.29	3.01	4.86	−0.13	−0.32	0.28	4.11	4.11	16.67	16.67	79.17	100.00	100.00
WTPD	−0.10	−0.32	0.21	3.01	4.86	−0.09	−0.26	0.22	4.09	4.09	20.83	16.67	75.00	100.00	100.00
Sample of prices (drop last observation)															
WGEKS															
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.33

TABLE 2 (CONTINUED)

	Mean					Median					% With Positive Bias				
Method	σ: 0	2	5	10	15	0	2	5	10	15	0	2	5	10	15
Fisher	0.00	−0.07	−0.63	−0.88	−0.48	0.00	−0.10	−0.79	−0.44	−0.44	0.00	16.67	20.83	16.67	37.50
Törnqvist	−0.01	−0.01	−0.35	−0.80	−0.54	−0.01	−0.01	−0.42	−0.53	−0.53	8.33	12.50	16.67	12.50	33.33
Chained															
CES (Sato–Vartia)	−0.48	−0.72	−1.08	−1.46	−1.31	−0.51	−0.76	−1.22	−1.23	−1.23	20.83	16.67	37.50	37.50	41.67
Fisher	−0.48	−0.94	−2.51	−3.31	−2.40	−0.51	−0.99	−2.16	−1.13	−1.13	20.83	20.83	25.00	37.50	37.50
Törnqvist	−0.50	−0.76	−1.86	−3.02	−2.39	−0.54	−0.84	−2.04	−2.22	−2.22	20.83	16.67	29.17	37.50	37.50
Geary–Khamis	−0.10	−0.15	0.27	1.94	3.24	−0.10	−0.15	0.27	2.57	2.57	16.67	25.00	79.17	91.67	95.83
WTPD	−0.07	−0.13	0.26	1.96	3.28	−0.05	−0.13	0.31	2.51	2.51	25.00	20.83	79.17	100.00	100.00
Sample of prices (drop first observation)															
WGEKS															
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	16.67
Fisher	0.00	−0.17	−1.26	−1.69	−1.18	0.00	−0.15	−1.08	−1.18	−1.18	0.00	0.00	12.50	4.17	25.00
Törnqvist	−0.01	−0.03	−0.66	−1.41	−1.12	−0.01	−0.02	−0.57	−1.07	−1.07	0.00	0.00	8.33	4.17	29.17
Chained															
CES (Sato–Vartia)	−0.97	−1.74	−4.25	−6.70	−6.94	−0.94	−1.74	−3.63	−5.70	−5.70	20.83	12.50	8.33	12.50	20.83
Fisher	−0.97	−2.20	−7.25	−9.99	−9.61	−0.94	−2.12	−5.19	−8.49	−8.49	20.83	8.33	4.17	12.50	16.67
Törnqvist	−1.01	−1.82	−5.91	−9.28	−9.26	−0.99	−1.80	−4.63	−7.48	−7.48	20.83	12.50	4.17	8.33	16.67
Geary–Khamis	−0.18	−0.37	0.28	3.17	5.19	−0.12	−0.33	0.29	4.12	4.12	12.50	12.50	70.83	100.00	100.00
WTPD	−0.08	−0.33	0.20	3.16	5.19	−0.07	−0.28	0.21	4.02	4.02	25.00	12.50	66.67	100.00	100.00

Notes: The table shows the annual average difference in percentage points from CES-WGEKS. This is averaged across products and cities.

computations, the i represents the price of a product for a particular store-product combination. This is to ensure that measured price change reflects actual changes in prices paid rather than changes in consumption patterns across stores who may offer different shopping experiences (see Ivancic and Fox (2013) for a comprehensive discussion of aggregation methods for scanner data).

We use these data in two main ways. First, we construct a panel data set of prices for 1,000 products for each of the cities and product categories. Second, we take a random sample of 5,000 products for each category and city.

The second data set is relatively self-explanatory. However, we note that, because we take a sample of actual prices, there will be a degree of missingness in the data. This is because products appear and disappear in the actual data.

In the panel data set, we use the scanner data to construct an artificial set of prices that are observed every period. We do this by first randomly selecting the prices of 1,000 products in certain stores. Each of these products is followed until they disappear. As each product disappears, we randomly select another product-store price and link this onto the price for the disappearing product.

Let us consider an example of how the panel data set is constructed. Suppose we are constructing price i' in the panel data set. We start by randomly selecting the price for product k from the full data set. Suppose product k is observed for seven periods; $p_{kt}, p_{kt+1}, \ldots, p_{kt+6}$. We actually exclude the first and last prices from use in creating the balanced panel—the reasons for doing so will become clearer when we outline our results in the following sections. In this case we define $p_{i'1} = p_{kt+1}, p_{i'2} = p_{kt+2}, \ldots, p_{i'5} = p_{kt+5}$. Price k disappears in period 7 and we do not use p_{kt+6}, so we need to find another price to link on to construct $p_{i'6}$. We randomly select a new product, say j, then we link this price as follows: $p_{i'6} = p_{i'5} \times \left(\frac{p_{jv+2}}{p_{jv+1}}\right)$. Here v is the first period that product j is observed in the market. Subsequent prices are calculated as $p_{i'7} = p_{i'6} \times \left(\frac{p_{jv+3}}{p_{jv+2}}\right)$ and so forth.[10] This process is continued until we have a data set that has 1,000 products observed over 144 time periods (12 years × 12 months).[11]

4. Results

We undertake a number of simulations using each of the product categories, cities, and the two data sets—sample and panel—discussed earlier. We do this for what we regard as a "plausible range" of values of σ for this type of data from $\sigma = 0$ to $\sigma = 15$. Our choice of this range reflected previous, and particularly the more recent, research in this area such as Broda and Weinstein (2010). They estimate various CES functions on barcode-level scanner data and find that the median σ is around 7–10.

[10]This approach is called the overlap pricing method. It is a standard approach for linking new products into an index such that any quality differences between the products do influence price change (ILO, 2004).

[11]Summary statistics on these data sets are given in the Online Appendix. In general, the average price and the standard deviation of prices in both data sets are similar to that in the complete data.

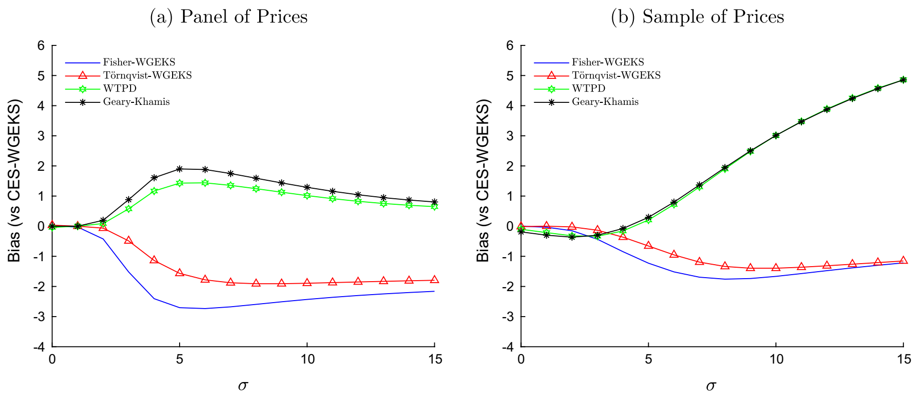

Figure 1. Index Bias

Note: The figures plot the annual average difference in percentage points from CES-WGEKS. This is averaged across products and cities. [Colour figure can be viewed at wileyonlinelibrary.com]

Our results are too voluminous to present in their entirety. Instead, we present average results for each type of data and index approach. The average results are constructed by equally weighing the results for each of the 24 product–city combinations. Table 2 presents the results for $\sigma = 0, 2, 5, 10, 15$ for each multilateral index number method that we implemented; (1) various WGEKS-based approaches using different bilateral index number formulae and weighting mechanisms, (2) chained indexes, (3) WTPD, and (4) the temporal GK approach. In addition, in the case of the panel data set, we present the results from constructing direct Fisher and Törnqvist indexes.

In each case, the table presents the percentage point difference in the index methods from using the WGEKS approach with the exact CES (Sato–Vartia) index as the bilateral comparator. The differences are presented in annual average terms, and we take the mean and median across cities and product categories. The final part of the table denotes the percentage of the 24 product–city combinations that exhibit a positive bias in each case. Figure 1 plots the bias for certain indexes for different σ.

4.1. *WGEKS and Direct Methods*

Let us begin by focusing on differences among the various WGEKS approaches that we consider. The CES (Sato–Vartia) WGEKS approach is our reference point, and therefore the bias for this method is always zero.

Turning first to the panel of prices data—as noted earlier this has no missing prices. The Fisher and Törnqvist WGEKS indexes reveal very little bias for low values of σ, from 0 to 2. This reflects the fact that the Törnqvist index is exact when $\sigma = 1$ and the Fisher index is exact when $\sigma = 0$. However, as σ rises the differences between the Fisher and Törnqvist WGEKS indexes, and the reference index becomes nontrivial. For $\sigma = 6$, the Fisher-WGEKS approach exhibits its largest level of bias of -2.73 percentage points per annum on average. For the Törnqvist-WGEKS approach, bias is largest at $\sigma = 9$ and equals -1.91 percentage points.

For $\sigma = 10$, the downward bias is large but decreasing. It equals −2.43 and −1.89 percentage points, respectively. At $\sigma = 15$, the bias is smaller still.

In the sample of prices data set, a downward bias in Fisher-WGEKS and Törnqvist-WGEKS indexes is also readily apparent. However, it is somewhat attenuated when compared with that observed in the panel data sets. In the sample data set, the bias in the Fisher-WGEKS approach is largest when $\sigma = 8$ and is equal to −1.75 percentage points per annum. For the Törnqvist-WGEKS, the bias is largest at $\sigma = 9$ and equals −1.40 percentage points per annum.[12]

The differences between the exact CES index, on one hand, and the Fisher and Törnqvist WGEKS, on the other hand, are certainly nontrivial. This has not been widely documented before, and it is not straightforward to find an explanation for this phenomenon. However, we note that Hill (2006) observed quite significant differences in bilateral index numbers for the quadratic-mean-of-order-r index, equation (15), for different values of r. He noted that for his data the spread between the maximum and minimum quadratic-mean-of-order-r indexes for different r often exceeded the Paasche–Laspeyres spread. This documents the fact that significant differences can arise, as we have found, between index methods, which provide similarly flexible approximations to consumer preferences.

In the case of the panel-of-prices data, we can calculate direct indexes. This is not a sensible approach in the sample-of-prices data because products appear and disappear. This means that the set of products matched with those sold in the first period will degrade significantly as the gap between the periods lengthens.

The direct Fisher and Törnqvist indexes also deviate from the exact CES index. The bias is somewhat smaller than that for the WGEKS approaches for values of σ below 5. However, for larger values of the elasticity of substitution, the bias is positive and significant. This contrasts the bias for the WGEKS indexes, which was negative. In the case of the Fisher index the bias peaks at 2.58 percentage points, whereas for Törnqvist it is 2.28 percentage points, for $\sigma = 5$.

Our results for the WGEKS and direct methods are somewhat unexpected. Our expectation was that these methods should closely approximate the exact CES results. It is surprising then that the differences are quite large. These results pose interesting questions about the source of these differences, these issues require further investigation.[13]

4.2. *Weighted TPD*

For the WTPD method, there are also significant differences from our reference point—the CES-WGEKS method. In the case of the panel of prices data set, this bias is fairly mild up to around $\sigma = 2$. For higher levels of σ, the bias rises and peaks around $\sigma = 6$ at 1.44 percentage points per annum. After this point, the bias falls, although remains positive and nontrivially large. These results appear fairly

[12]In the Online Appendix, we provide figures that outline the level of bias for each product and city combination. These figures show that each method does tend to have a consistent bias across the various products and cities.

[13]The differences between the CES and superlative indexes are not easy to identify based on algebraic manipulations. Instead, they appear to relate to numerical issues and how the different formulae behave given asymmetries in the distribution of prices and price relatives. This is an interesting line of inquiry that warrants further investigation.

consistent with our theoretical expectations. We noted that the WTPD method can be justified based on $\sigma = 1$ or an extremely large value of σ. Our results confirm this but also point to the prevalence of positive bias through the intermediate range and a turning point around $\sigma = 6$.

Interestingly, the results are quite different for the sample data set. Here the WTPD method exhibits minor levels of bias up to around $\sigma = 5$. However, after this time the level of bias is positive and increases rapidly. At $\sigma = 10$, the bias is a whopping 3.01 percentage points per annum, and it rises to 4.86 percentage points at $\sigma = 15$. The differences in the results for WTPD from the panel and sample data sets imply that this method is strongly influenced by new and disappearing goods.

4.3. *The Geary–Khamis Method*

Now, turning to the temporal GK method. The results are broadly similar to those for WTPD. In the panel of prices the bias, relative to the CES WGEKS approach, is mild for small values of σ. It rises and peaks at 1.90 percentage points per annum for $\sigma = 5$ and then decreases for larger values of σ.

However, in the case of the sample data set, the bias is positive and much more significant. As we saw for the WTPD method, the bias starts rising rapidly after around $\sigma = 5$ and reaches 3.01 percentage points at $\sigma = 10$ and 4.86 percentage points at $\sigma = 15$. This also points to a significant interaction between the GK method and new and disappearing goods.

4.4. *Chained Indexes*

Perhaps the most surprising results are the chain drift that is evident for the CES index in the sample of prices data set. In the panel data set, with no missingness, the chained CES indexes are transitive and therefore have no bias. However, in the sample of prices data set, a significant divergence can be observed between the chained CES indexes and the CES-WGEKS indexes. The size of this drift increases as σ becomes larger. Significant chain drift is also evident for the Fisher and Törnqvist chained indexes.

This result is startling. As noted in the introduction, the standard explanation for chain drift is that it relates to stockpiling (Ivancic *et al.*, 2011; de Haan and van der Grient, 2011). However, we have simulated product expenditure shares using a classical demand framework. This excludes the possibility of stockpiling by construction. What then is the explanation for such significant chain drift? We turn to the impact that life cycle pricing has on chained price indexes, and related issues, in the next section.

5. Life Cycle Pricing and Index Bias

One puzzle posed by our simulation results is the high degree of chain drift in the sample of prices indexes. This leads to a large difference between the chained and WGEKS indexes even though the expenditure shares have been constructed based on the CES demand system, where stockpiling plays no role. We argue that this is related to life cycle pricing phenomena. More specifically, in many product categories run-out sales—associated with big falls in price—are observed at the end of a product's life. These price falls are not balanced by commensurate price rises elsewhere. It is straightforward to show that this can lead to a significant downward bias in the resulting chained indexes. Let us explore this algebraically as well as using our data.

To see algebraically how run-out sales can lead to downward chain drift, consider an example where we have T periods, $t = 1, 2, \ldots, T$. Let us suppose that the Törnqvist bilateral index is used in constructing both the chained and multilateral GEKS indexes. The Törnqvist index and GEKS approach are used for clarity. However, the principle applies more generally were a different index to be used, such as the Sato–Vartia or Fisher indexes, and even for other multilateral methods such as GK and TPD.[14] The chained and GEKS multilateral indexes, between the first and last periods, can be written, respectively, as follows:

$$P_{1T}^{T(C)} = P_{12}^{T} \times P_{23}^{T} \times \cdots \times P_{T-1T}^{T} \tag{19}$$

$$P_{1T}^{T(GEKS)} = \left(\frac{\prod_{t=1}^{T} P_{tT}^{T}}{\prod_{t=1}^{T} P_{t1}^{T}} \right)^{\frac{1}{T}}. \tag{20}$$

To illustrate the differences between these two methods, let us consider a highly stylized situation where all prices are unchanged except for those of a single good, say good j. Suppose this product is sold in period 1, goes on sale in period 2, and then disappears from the market in all subsequent periods. In this case the chained index is equal to

$$P_{1T}^{T(C)} = \left(\frac{p_{j2}}{p_{j1}} \right)^{\frac{\tilde{s}_{i1} + \tilde{s}_{i2}}{2}}. \tag{21}$$

In comparison, for the Törnqvist-GEKS index, we have

$$P_{1T}^{T(GEKS)} = \left(\frac{p_{j2}}{p_{j1}} \right)^{\frac{\tilde{s}_{i1} + \tilde{s}_{i2}}{2T}}. \tag{22}$$

The difference is stark. In the chained case, the price fall gets a significantly higher weight than in the multilateral case. In fact, the multilateral case gives the price change, $\frac{p_{j2}}{p_{j1}}$, just $1/T$th the weight given to the same price relative in the chained index. The lower weight for the good j in our simple example reflects the fact that this good was only sold in periods 1 and 2 and so it can only be used in bilateral comparisons between these two periods. In the case of the multilateral indexes, it

[14]These effects can also occur at the elementary index level when a sample of prices is used, products disappear over time, and there are life cycle trends in prices (Melser and Syed, 2016, 2017).

does not appear in any other bilateral comparisons; therefore, if T is large it will be missing in the vast majority of cases. This reduces its ability to influence measured price change in the multilateral case. This emphasizes an important property of multilateral indexes relative to chained indexes; multilateral indexes tend to reduce the influence of new and disappearing goods because these products appear in fewer bilateral comparisons.

Another interesting finding from our results was the rising upward bias in WTPD and GK in the sample of prices data set as σ rose. In the case of the panel of prices, the bias tended to fall—after a certain point—as σ rose. It seems likely that the behavior of WTPD and GK in the sample data set is also related to the effects of new and disappearing goods.

Given the apparent importance of life cycle pricing, we undertake some further empirical investigations to gauge its influence.

5.1. *Quantifying the Effects of Life Cycle Pricing*

For life cycle pricing effects to be responsible for the chain drift we observe in our simulated indexes, life cycle pricing effects must exist. We investigate this by running a WTPD-type regression on the full data set. We estimate a model such as (12). However, we augment this model with a dummy variable for whether it was the last period that a product was observed to sell at a particular store. In a second regression, we include a dummy variable for whether it was the first sale of a product in a store. The results for the "disappearing" and "new" dummy variables, for each product category and city, are presented in Table 3.

Our results indicate a clear tendency for products to be subject to run-out sales. However, there are limited life cycle effects on prices in the introductory period. In the first panel of Table 3, we see a strong negative effect on price in the last period in which they are observed. In fact, there are only 2 cases of a significant positive effect, 4 insignificant coefficients, and the remaining 18 coefficients are significantly negative. On average, across all products and markets, prices are around 3 percent lower in their last period. Turning to price levels in the first period that an item is sold, we see much less significant differences. Fewer coefficients are statistically significant, and there is a mix of positively and negatively signed coefficients. The overall average indicates very little difference in price levels for new products in the first period in which they are sold in a given store. Thus most of the life cycle pricing dynamics appear to be run-out sales rather than introductory specials.

To explore the effect of product life cycle dynamics on chain drift, we undertake some further simulations. First, we estimate the various indexes on the sample of prices data set, but we drop the price data for a product in the last period it was sold. For completeness, we run a further simulation but this time dropping the price for the first period a product was sold. The results are presented in the bottom half of Table 2 and are plotted in Figures 2 and 3.

The results of this exercise are interesting. They show that the level of chaining bias is significantly attenuated when last prices are dropped but little changed—even worsened—when first prices are dropped. For the CES index, the chaining bias for $\sigma = 5$ falls (in absolute value) from −4.04 to −1.08 percentage points when

TABLE 3
Price Change for New and Disappearing Goods (Full Data Set)

	Coefficient on Last Price			Average
	Dallas	Los Angeles	New York	
Beer	−0.0008	−0.0090***	0.0062***	−0.0012
Carbonated beverages	0.0218***	0.0008	−0.0022	0.0068
Coffee	−0.0542***	−0.0388***	−0.0244***	−0.0391
Margarine and butter	−0.0209***	−0.0027	−0.0096***	−0.0111
Peanut butter	−0.0242***	−0.0238***	−0.0238***	−0.0239
Soup	−0.0731***	−0.0791***	−0.0308***	−0.0610
Toilet tissue	−0.0167***	−0.0111***	0.0034	−0.0082
Toothpaste	−0.1140***	−0.1207***	−0.0356***	−0.0901
Average	−0.0353	−0.0356	−0.0146	−0.0285

	Coefficient on First Price			
	Dallas	Los Angeles	New York	Average
Beer	0.0014	−0.0002	0.0013	0.0008
Carbonated beverages	−0.0063***	−0.0018	0.0004	−0.0026
Coffee	−0.0033*	0.0031*	−0.0014	−0.0005
Margarine and butter	−0.0051*	0.0096***	−0.0012	0.0011
Peanut butter	0.0079***	0.0035	−0.0073***	0.0014
Soup	0.0049***	0.0046***	0.0126***	0.0074
Toilet tissue	−0.0015	−0.0034**	0.0026	−0.0008
Toothpaste	−0.0006	−0.0097***	0.0021	−0.0028
Average	−0.0003	0.0007	0.0011	0.0005

***p-value < 0.01, **p-value < 0.05, *p-value < 0.1.

we drop last prices. There are commensurate falls in the chained Fisher and Törnqvist indexes with the latter plotted in Figure 2. The results, illustrated in Figure 3, also show a significantly reduced bias in the WTPD and GK methods when the last price is dropped.[15]

The results of our simulation are quite clear in pinpointing the importance of run-out sales in generating chain drift. However, it is useful to take this a step further and try to more accurately identify the amount of the chain drift that is due to life cycle pricing behavior. We do this by calculating the indexes above using the full data. That is, instead of generating expenditure shares and quantities using the

[15]As pointed out by a referee, another way of dealing with the run-out sales problem would be to use the assortment adjustment approach of Feenstra (1994). If products receive a disproportionate share of expenditure in their last period, as is likely if they are subject to run-out sales and the elasticity of substitution is greater than one, then applying the Feenstra adjustment would tend to raise the cost of living in the subsequent period. The Feenstra adjustment is based on a comparison of the expenditure shares of new and disappearing goods—if the share on disappearing goods is higher than that on new goods, this will tend to raise the cost of living. This adjustment would counterbalance the decreases in price recorded during run-out sales. An important area of future research is how the various multilateral approaches perform in terms of approximating both the matched price change and the welfare effects of changing assortments reflected in the approach of Feenstra (1994).

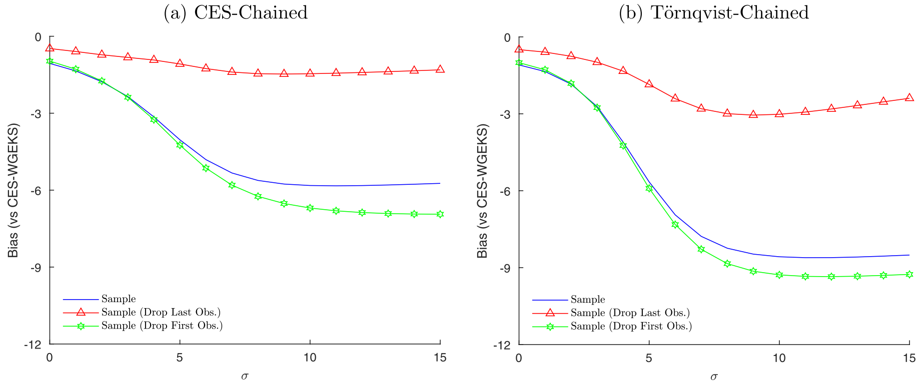

Figure 2. Index Bias—Chain Drift (Sample of Prices)
Note: The figures plot the annual average difference in percentage points from CES-WGEKS. This is averaged across products and cities. [Colour figure can be viewed at wileyonlinelibrary.com]

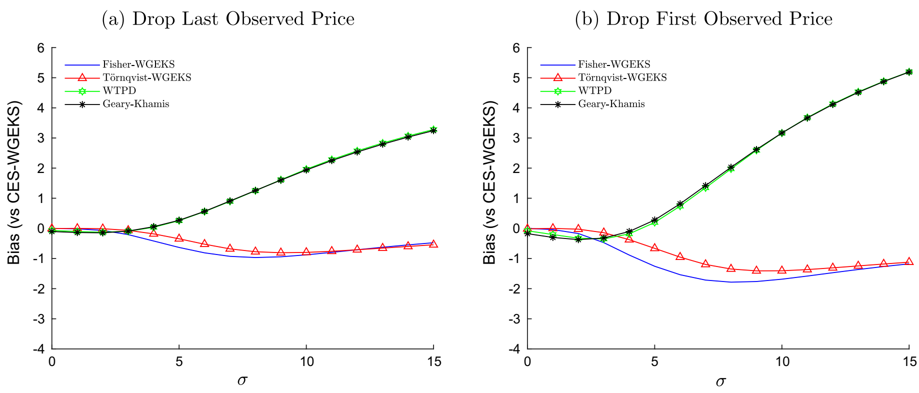

Figure 3. Index Bias—Dropping First and Last Observed Prices (Sample of Prices)
Note: The figures plot the annual average difference in percentage points from CES-WGEKS. This is averaged across products and cities. [Colour figure can be viewed at wileyonlinelibrary.com]

CES cost function, we use the actual quantities and shares. This also addresses a possible issue from using simulated expenditure shares in the presence of run-out sales. Although it may be the case that products are sold at very low prices in their last period, it may also be that limited quantities are available at these prices. The use of generated expenditure shares may then exaggerate the quantities that are bought at these prices. Using the actual prices and quantities addresses this issue. It also allows us to get some idea of the relative importance of life cycle pricing compared with stockpiling in generating chain drift, although we note that these two effects are likely to interact and reinforce each other. For example, consumers may use run-out sales as an opportunity to stockpile a product.

In Table 4, we undertake estimation for three different cases using the full data set: (1) using all observations, (2) dropping last prices, and (3) dropping first prices. The results broadly confirm those of the simulation. Let us focus on the chained Törnqvist index. Using all observations, the chain drift is estimated at −0.75 percentage points on average per annum. If we drop the last observed price, then

TABLE 4
INDEX BIAS AND LIFE CYCLE PRICING (FULL DATA SET)

	All Obs.			Drop Last Obs.			Drop First Obs.		
	Mean	Median	% Positive	Mean	Median	% Positive	Mean	Median	% Positive
WGEKS									
CES (Sato–Vartia)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fisher	−0.01	−0.01	33.33	0.00	−0.02	33.33	0.00	−0.01	37.50
Törnqvist	−0.02	−0.02	20.83	−0.01	−0.01	25.00	−0.02	−0.01	25.00
Chained									
CES (Sato–Vartia)	−0.41	−0.32	41.67	−0.09	−0.11	41.67	−0.40	−0.29	41.67
Fisher	−0.99	−0.82	20.83	−0.63	−0.61	25.00	−0.96	−0.80	20.83
Törnqvist	−0.75	−0.66	20.83	−0.38	−0.29	25.00	−0.72	−0.54	20.83
Geary–Khamis	−0.09	−0.03	37.50	−0.05	−0.01	45.83	−0.10	−0.02	37.50
WTPD	−0.05	0.00	54.17	−0.02	0.01	54.17	−0.05	0.00	50.00

Notes: The table shows the annual average difference in percentage points from CES(Sato–Vartia)-WGEKS. This is averaged across products and cities.

the mean chain drift falls (in absolute value) to −0.38 percentage points. This is a sizable reduction in chain drift of around half. The relative size of the decrease is slightly smaller, but still sizable at around one-third for the chained Fisher. For the chained CES (Sato–Vartia) index, run-out sales appear to eliminate around three-quarters of the bias. The WTPD and GK also exhibit reduced bias when the last price is dropped. These results, coupled with those of the simulation, provide clear evidence of the strong influence that life cycle pricing plays in inducing chain drift and in introducing a degree of bias in the WTPD and GK methods.

6. Conclusion

The purpose of this paper has been to examine the properties of some of the "new" methods used to calculate temporal price indexes on scanner data. These methods are now becoming more widely used, but there is a deficit in our understanding of their properties. Two issues are particularly important in this regard: first, the extent to which the various methods may embody substitution bias and second, how the methods perform in the presence of new and disappearing goods. To examine these questions, we built on the approach of Diewert and Fox (2018) in developing our simulation framework. Our simulations used actual price data, sourced from the IRI scanner data set (Bronnenberg *et al.*, 2008), but created expenditures/quantities based on CES preferences. The results yielded several interesting findings.

First, for the panel data set with no missing prices, there was a surprising difference between the exact CES index and the Fisher and Törnqvist WGEKS and direct indexes. On average, the Fisher and Törnqvist WGEKS indexes tend to understate price change, whereas the direct indexes overstated price change. These findings are interesting and warrant further research.

Second, we found that the WTPD and GK methods tend to overstate price change. This was the case in the panel data set but to a much greater extent in the sample of prices where products appeared and disappeared. These results imply that a degree of caution should be exercised in using these approaches.

Third, our results indicate that a significant portion of the chain drift that is observed in indexes using scanner data is likely to be the result of life cycle pricing phenomena. This contrasts with conventional wisdom, which is that chain drift is primarily because of stockpiling by consumers. In particular, much of the downward drift in our chained indexes arises because of the commonly observed practice of run-out sales as products exit the market. We show that these run-out sales get much higher weight in chained indexes than in multilateral indexes, and that this can lead to downwards drift. In our simulations, and estimation on the full data set, we find that removing the last price of a product—often a sale price—significantly attenuates chain drift. It also ameliorates some of the bias evident in the WTPD and GK methods.

References

ABS, "Making Greater Use of Transactions Data to Compile the Consumer Price Index, Australia," Information Paper No. 6401.0.60.003, 2016.

______, "An Implementation Plan to Maximise the Use of Transactions Data in the CPI," Information Paper No. 6401.0.60.004, 2017.
Aizcorbe, A., C. Corrado, and M. Doms, "When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures," Working Paper No. 2003–14, Federal Reserve Bank of San Francisco, 2003.
Arrow, K. J., H. B. Chenery, B. S. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Review of Economics and Statistics*, 43(3), 225–50, 1961.
Balk, B. M., "A Simple Method for Constructing Price Indices for Seasonal Commodities," *Statistische Hefte*, 22(1), 1–8, 1981.
Broda, C. and D. E. Weinstein, "Product Creation and Destruction: Evidence and Price Implications," *American Economic Review*, 100(3), 691–723, 2010.
Bronnenberg, B. J., M. Kruger, andC. F. Mela, "The IRI Marketing Data Set," *Marketing Science*, 27(4), 745–8, 2008.
Caves, D. W., L. R. Christensen, and W. E. Diewert, "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity," *Econometrica*, 50(11), 1393–414, 1982.
Chessa, A. G., "A New Methodology for Processing Scanner Data in the Dutch CPI," *European Review of National Accounts and Macroeconomic Indicators*, 1, 49–69, 2016.
de Haan, J., "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches," *Paper presented at the Seventh EMG Workshop*, UNSW, Sydney, Australia, 2007.
______ "Pondering Over the CES Price Index," *Presentation at 2019 EMG Workshop*, UNSW, Sydney, 2019.
de Haan, J. and H. A. van der Grient, "Eliminating Chain Drift in Price Indexes Based on Scanner Data," *Journal of Econometrics*, 161(1), 36–46, 2011.
Diewert, W. E., "Exact and Superlative Index Numbers," *Journal of Econometrics*, 4(2), 115–45, 1976.
______, "Axiomatic and Economic Approaches to International Comparisons," in A. Heston and R. E. Lipsey (eds), *International and Interarea Comparisons of Income, Output and Prices, Studies in Income and Wealth*, Vol. 61, The University of Chicago Press, Chicago, 13–87, 1999.
______, "Weighted Country Product Dummy Variable Regressions and Index Number Formulae," *Review of Income and Wealth*, 51(4), 561–71, 2005.
Diewert, W. E. and K. J. Fox, "Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data," Discussion Papers 2018–13, School of Economics, The University of New South Wales, 2018.
Eltetö, O. and P. Köves, "On a Problem of Index Number Computation Relating to International Comparison," *Statisztikai Szemle*, 42, 507–18, 1964. [in Hungarian]
Feenstra, R., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84(1), 157–77, 1994.
Feenstra, R. C., R. Inklaar, and M. P. Timmer, "The Next Generation of the Penn World Table," *American Economic Review*, 105(10), 3150–82, 2015.
Geary, R. G., "A Note on Comparisons of Exchange Rates and Purchasing Power between Countries," *Journal of the Royal Statistical Society Series A*, 121(1), 97–9, 1958.
Gini, C., "On the Circular Test of Index Numbers," *Metron*, 9(9), 3–24, 1931.
Hill, R. J., "Superlative Index Numbers: Not All of Them Are Super," *Journal of Econometrics*, 130(1), 25–43, 2006.
ILO, *Consumer Price Index Manual: Theory and Practice*, Produced by: ILO/IMF/OECD/UNECE/Eurostat/The World Bank, Geneva, 2004.
Ivancic, L., W. E. Diewert, and K. J. Fox, "Scanner Data, Time Aggregation and the Construction of Price Indexes," *Journal of Econometrics*, 161(1), 24–35, 2011.
Ivancic, L. and K. J. Fox, "Understanding Price Variation Across Stores and Supermarket Chains: Some Implications for CPI Aggregation Methods," *Review of Income and Wealth*, 59(4), 629–47, 2013.
Khamis, S. H., "A New System of Index Numbers for National and International Purposes," *Journal of the Royal Statistical Society Series A*, 135(1), 96–121, 1972.
Kokoski, M. F., B. R. Moulton, and K. D. Zieschang, "Interarea Price Comparisons for Heterogeneous Goods and Several Levels of Commodity Aggregation," in A. Heston and R. E. Lipsey (eds), *International and Interarea Comparisons of Income, Output and Prices, from Studies in Income and Wealth*, Vol. 61, The University of Chicago Press, Chicago, 123–69, 1999.
Kruger, M. W., "Research Use of the IRI Marketing Data Set: Bibliography," Discussion Paper, 2017.
Krsinich, F., "The FEWS Index: Fixed Effects with a Window Splice," *Journal of Official Statistics*, 32(2), 375–404, 2016.
Melser, D., "Accounting For The Effects Of New And Disappearing Goods Using Scanner Data," *Review of Income and Wealth*, 52(4), 547–68, 2006.

Melser, D. and I. A. Syed, "Life Cycle Price Trends and Product Replacement: Implications for the Measurement of Inflation," *Review of Income and Wealth*, 62(3), 509–33, 2016.
______, "The Product Life Cycle and Sample Representativity Bias in Price Indexes," *Applied Economics*, 49(6), 573–86, 2017.
Melser, D., "Scanner Data Price Indexes: Addressing Some Unresolved Issues," *Journal of Business and Economic Statistics*, 36(3), 516–22, 2018.
______, "Valuing the Quantity and Quality of Product Variety to Consumers," *Empirical Economics*, 57(6), 2107–28, 2019.
Rao, D. S. P. and K. S. Baneerjee, "A Multilateral System of Index Numbers Based on the Factorial Approach," *Statistiche Hefte*, 27, 297–312, 1986.
Rao, D. S. P. and M. P. Timmer, "Purchasing Power Parities for Industry Comparisons Using Weighted Elteto-Koves-Szulc (EKS) Methods," *Review of Income and Wealth*, 49(4), 491–511, 2003.
Rao, D. S. P., "On the Equivalence of Weighted Country-Product-Dummy (CPD) Method and the Rao-System For Multilateral Price Comparisons," *Review of Income and Wealth*, 51(4), 571–80, 2005.
Sato, K., "The Ideal Log-Change Index Number," *Review of Economics and Statistics*, 58(2), 223–28, 1976.
Summers, R., "International Price Comparisons Based upon Incomplete Data," *Review of Income and Wealth*, 19(1), 1–16, 1973.
Szulc, B., "Indices for Multiregional Comparisons," *Przeglad Statystyczny*, 3, 239–54, 1964. [in Polish]
Vartia, Y., "Ideal Log-Change Index Numbers," *Scandinavian Journal of Statistics*, 3(3), 121–26, 1976.
von Auer, L., "The Generalized Unit Value Index Family," *Review of Income and Wealth*, 60(4), 843–61, 2014.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Table A.1: Scanner Data Summary Statistics (Panel Dataset)
Table A.2: Scanner Data Summary Statistics (Sample Dataset)
Figure B.1: Bias for Panel of Prices
Figure B.2: Bias for Sample of Prices