IARIW 2021

IARIW 2021

Monday 23 – Friday 27 August

Statistisk sentralbyrå
Statistics Norway

Econometric Interpretations of Index Numbers

Adam Gorajek

(Reserve Bank of Australia)

Paper prepared for the 36th IARIW Virtual General Conference
August 23-27, 2021
Session 12: Comparing Income and Price Levels and International Income Inequality
Time: Wednesday, August 25, 2021 [14:00-16:00 CEST]

Econometric Interpretations of Index Numbers

Adam Gorajek*

Reserve Bank of Australia; gorajeka@rba.gov.au

August 13, 2021

I translate the task of measuring price indexes into the language of econometrics, showing that the task amounts to estimating fixed effects in a simple model of quality adjustment. Earlier translations are less general and suffer from a misspecification relating to product weighting. I then use the new translation to argue three points. First, the "stochastic approach" to choosing index functions is unhelpful for choosing index functions, because in its complete form it can justify all of them. Second, the same feature of the stochastic approach makes it possible to calculate confidence intervals for any index type. Third, the literature uses flawed arguments for swapping the time-dummy hedonic method of quality adjustment with hedonic imputation.

Key words: quality adjustment, index function, stochastic approach, confidence intervals, hedonic regression

JEL codes: C18, C43, C51

*For their input I thank Bert Balk, Anthony Brassil, Michele De Nadai, Erwin Diewert, Jan de Haan, Gholamreza Hajargasht (discussant), Denzil Fiebig, Kevin Fox, Robert Hill, SeoJeong (Jay) Lee, Alicia Rambaldi, Prasada Rao, John Simon, Franklin Soriano (discussant), Michael Webster, Grahame White, three anonymous referees, and participants at several conferences and seminars. Rosie Wisbey sourced some elusive references. Paula Drew and Sonja Heijn gave editorial assistance. I also thank the James M Kilts Center (University of Chicago Booth School of Business) for data, and acknowledge an Australian Government Research Training Scholarship. The views in this paper are mine and not necessarily those of the Reserve Bank of Australia.

1 Introduction

In principle, macroeconomic price indexes exclude the effects of quality change, which then rightfully appear in indexes of real activity. For example, improvements in product quality are meant to show up in real output growth, not the output deflator. But accounting for quality change is difficult, and prevailing solutions are incomplete. The resulting data distortions make it harder for policymakers to detect genuine economic changes and to identify legitimate economic relationships. The collective effort to solve this problem is now over a century old.

To help, I translate the task of measuring price indexes into the language of econometrics. The translation offers new price index interpretations, which I argue reveal problems and opportunities in economic measurement.

Translations similar to mine already appear in the price index literature, and two parts in particular. First, when product types are static (but their market shares still move) the main tools for handling quality change are index functions. The "stochastic approach" to choosing index functions, in its existing form, already distinguishes some of those functions by how well they estimate parameters in econometric descriptions of the price index measurement problem (see Selvanathan & Rao 1994, Diewert 2010, and Rao & Hajargasht 2016). Second, when product types are dynamic, practitioners sometimes account for quality change using the method of "time-dummy hedonic" regression. That too targets parameters in econometric descriptions of the price index measurement problem (see Diewert 2005a and Reis & Santos Silva 2006).

My translation builds on these earlier ones by generalising and unifying them. It also corrects a pervasive source of model misspecification relating to product weighting—a misspecification that has fuelled unnecessary doubt about the relevance of econometric perspectives to economic measurement (in Diewert 2005a and Diewert 2010, for example). The form of my translation is a model of quality adjustment that reduces the task of price index measurement to answering three conceptual questions:

1. What is the true population of interest?
2. What is the right measure of product quality?
3. What is the appropriate metric for central tendency?

I show that prominent index functions and time-dummy hedonic methods can all be understood as corresponding to different answers to these three questions and, thus, to special cases of the quality adjustment model. Contrary to the prevailing view in the price index literature, sensible econometrics can justify more than just a few index functions.

What, then, are the best answers to these three central questions? Finding out would reveal the best index functions to use, but I fail to offer answers in this paper. One could back out answers using the index functions recommended by the selection criteria of the so-called economic or axiomatic approaches.[1] But that would presuppose the functions one wants to discover; it would not reveal anything new. The bottom line here is that unless scholars can find the best answers to the three central questions without relying on the economic or axiomatic approaches, the stochastic approach to choosing index functions does not actually help choose index functions. Left to its own, and in a more complete form (such as my translation), the stochastic approach can justify all of them.

This problem with the stochastic approach also opens an opportunity. In particular, scholars of the approach have long argued that it offers the best framework for estimating index

[1]The economic approach distinguishes functions by how well they measure the changing cost of attaining a given economic objective, such as an amount of output or living standard (Diewert 1981 is a review). The axiomatic approach—also called the "test" or "instrumental" approach—distinguishes functions by their ability to satisfy desirable mathematical criteria (Balk 2008 is a review). I do not detail either approach in this paper.

uncertainty. For all index numbers that can be legitimately understood as econometric estimators, practitioners can use textbook econometric tools to produce confidence intervals or standard errors (as in Clements & Izan 1981, Clements & Izan 1987, Rao & Selvanathan 1992, and Rao & Hajargasht 2016). By showing that all price indexes are sensible econometric estimators, my translation reveals that one can always produce these uncertainty measures.

For a final application, I use the translation to reveal flaws in influential criticisms of the time-dummy hedonic method of quality adjustment, as argued by Berndt & Rappaport (2001), National Research Council (2002; this is a report solicited by the US Bureau of Labor Statistics), Pakes (2003), Reis & Santos Silva (2006), and Diewert et al. (2009). I explain that, in effect, all of these sources criticise the time-dummy hedonic method because it restricts the implicit measure of product quality to be constant across the time periods in the hedonic regression, even if the time periods are just two adjacent ones. And yet my translation shows that time-constant measures of quality are also implicit features of all index functions that statistical agencies use today.[2] So unless the criticism is intended to cover index functions as well, time-constant quality is not a logically consistent basis on which to challenge the time-dummy hedonic method. Putting this another way, the use of time-constant quality makes the time-dummy hedonic method (for when product types are dynamic) cohere with index functions (for when product types are static). Moreover, there are conceptual reasons to favour the time-constant quality feature. Thus I challenge the push to move away from the time-dummy hedonic method. Reversing this trend matters because alternative methods—most notably hedonic imputation—often yield different economic narratives (Diewert et al. 2009).

Regarding the structure of this paper, to give context I first introduce the standard econo-

[2]Assuming constant quality across the time periods in a regression is not to assume constant quality across all time; dynamic definitions of quality still arise through the standard practice of chaining.

metric model of the measurement literature. The model already nests those most commonly used in the stochastic approach and time-dummy hedonic regression. I then correct and generalise the model, showing that the outcome unifies practically all bilateral and multilateral price index functions in the literature. That result is my econometric translation. The paper finishes with the applications.

2 The standard model generalises in three ways

2.1 The generalisation

The most common econometric model in the price index literature comes from Court (1939). In this paper I call it the "standard model". It describes prices in some product market using assumptions A1 to A3:

A1

$$ln(p_{it}) = \alpha_t + \boldsymbol{\beta' spec_i} + \varepsilon_{it} \qquad (t = 1, ..., T), \tag{1}$$

where p_{it} is the price of product i in time period or territory t, α_t is a fixed effect for each t, $\boldsymbol{\beta}$ is a vector of parameters, $\boldsymbol{spec_i}$ is a vector of (non-collinear) product specifications, ε_{it} is an error term, and $T \geq 2$.

A2 Each $(p_{it}, \boldsymbol{x_{it}})$ observation is a random sample from the population of time-product (or time-territory) pairs, where $\boldsymbol{x_{it}}$ is a vector of all the implied regressors in equation (1).

A3 The regressors are orthogonal to the errors. That is, $\mathbb{E}[\boldsymbol{x_{it}}\varepsilon_{it}] = \mathbf{0}$.

When working with this model, the practitioner usually aims to discover the differences between the α_t. For example, if t stands for time, $exp(\alpha_n - \alpha_m)$ measures quality-adjusted price growth from period $t = m$ to $t = n$. If t stands for territory, $exp(\alpha_n - \alpha_m)$ measures purchasing power. Going forward I will usually leave the territory option as implied.

Common special cases of the model differ along several other dimensions as well:

- The set of available product types can be dynamic or static. When the set is static, time becomes orthogonal to the regressors in $\boldsymbol{spec_i}$. Including $\boldsymbol{\beta' spec_i}$ is then irrelevant for defining the population price index, and the model takes the form used in the stochastic approach to index functions. When the set is dynamic, however, including $\boldsymbol{\beta' spec_i}$ can matter a lot. The model then takes the form used in time-dummy hedonic regression.[3]

- The types of regressors in $\boldsymbol{spec_i}$ can differ. Sometimes they are product attributes, in which case the model becomes "hedonic". At other times they are product dummies, in which case $\boldsymbol{\beta' spec_i}$ is a fixed effect, as in a so-called country-product-dummy model described by Diewert (2005b).

- The size of T can vary. Territory applications are often multilateral, so $T \geq 3$, as in the model versions that support official calculations of purchasing power parities (World Bank 2013). Time applications are often bilateral, so $T = 2$. Successive $e^{\alpha_2 - \alpha_1}$ then chain to form longer time series.

I propose a generalisation of the standard model that replaces A1 to A3 with A1' to A3'. In this paper I call it the "general model":

A1'

$$f\left(\frac{p_{it}}{quality_{it}}\right) = \alpha_t + \varepsilon_{it} \qquad (t = 1, ..., T), \tag{2}$$

where function $f(\cdot)$ and amounts $quality_{it}$ are new. I define and discuss them soon.

[3] Practitioners occasionally replace the log function in equation (1) with the identity function. Diewert (2005a) challenges that practice.

For now I just note that function $f(\cdot)$ must be invertible, with forms that are best not chosen arbitrarily.

A2' Each observation is a random sample from the population of time-product (or time-territory) pairs.

A3' For all t, the errors satisfy $\mathbb{E}_w[\varepsilon_{it}] = 0$. The w subscript in this new expectation operator $\mathbb{E}_w[\cdot]$ indicates that integration now occurs over a modified density function; the usual density function, in $\mathbb{E}[\cdot]$, has been multiplied throughout by a weighting expression $w_{it}/\mathbb{E}[w_{it}]$. I give a fuller explanation shortly.

These new assumptions imply that

$$\alpha_t = \mathbb{E}_w\left[f\left(\frac{p_{it}}{quality_{it}}\right)\right]. \tag{3}$$

The measurement task is now to discover the generalised target index

$$P_{m,n}^{general} \equiv \frac{f^{-1}(\alpha_n)}{f^{-1}(\alpha_m)} \tag{4}$$

$$= \frac{f^{-1}\left(\mathbb{E}_w\left[f\left(\frac{p_{in}}{quality_{in}}\right)\right]\right)}{f^{-1}\left(\mathbb{E}_w\left[f\left(\frac{p_{im}}{quality_{im}}\right)\right]\right)}. \tag{5}$$

I will show that just about all price index functions can be understood as estimators that are econometrically consistent for this generalised target index, and that it has a simple interpretation. What will differentiate each index are choices of $f(\cdot)$, measures of $quality_{it}$, and settings for w_{it}. For example, the target indexes defined by the standard model are the special cases for which $f(\cdot) = ln(\cdot)$, $quality_{it} = exp(\boldsymbol{\beta' spec_i})$, and $w_{it}/\mathbb{E}[w_{it}] = 1$ (which implies $\mathbb{E}_w[\cdot] = \mathbb{E}[\cdot]$). The general model allows everything that practitioners might already like about the standard one.

2.2 The role of $f(\cdot)$

$f(\cdot)$ is a function, continuous and strictly monotone over the support of $p_{it}/quality_{it}$. In equation (5) it defines the operator $f^{-1}(\mathbb{E}_w[f(\cdot)])$, a central tendency measure called a quasilinear mean. Writing the quasilinear mean as $\mathbb{M}_w^f[\cdot] \equiv f^{-1}(\mathbb{E}_w[f(\cdot)])$, the generalised target index in (5) shortens to

$$P_{m,n}^{general} \equiv \frac{\mathbb{M}_w^f\left[\frac{p_{in}}{quality_{in}}\right]}{\mathbb{M}_w^f\left[\frac{p_{im}}{quality_{im}}\right]}. \tag{6}$$

Several types of quasilinear mean are well known. For example, $f(\cdot) = (\cdot)$ defines an arithmetic mean, $f(\cdot) = ln(\cdot)$ defines a geometric mean, $f(\cdot) = (\cdot)^{-1}$ defines a harmonic mean, and $f(\cdot) = (\cdot)^2$ defines a quadratic mean. Each of these belongs to a more general class of quasilinear means called "power" means, whereby either $f(\cdot) = (\cdot)^\theta$ for non-zero θ, or $f(\cdot) = ln(\cdot)$. Power means are the only quasilinear means to satisfy linear homogeneity, which matters here because the target index in equation (6) becomes invariant to changes in the units used to measure product prices. Otherwise the target index would fail tests important to the axiomatic approach to choosing index functions (see International Labour Office et al. 2020, p.180, especially the proportionality test). Muliere & Parmigiani (1993) give an excellent summary of quasilinear mean properties.

Within the measurement literature, my use of quasilinear means looks most like Hajargasht & Rao (2019), which includes expressions resembling (5). A difference is that their work is deterministic, applying what are effectively sample versions of quasilinear means. So they do not unify hedonic regression with index functions. They also use less general versions of $quality_{it}$ and thus describe fewer index functions. Brachinger et al. (2018) use a similar expression to (5) as well, but without any role for objects like $quality_{it}$ or w_{it}. So they too describe fewer index functions. In a hedonic regression context, Reis & Santos Silva

(2006) discuss a transformation with role equivalent to $f(\cdot)$, using it to distinguish between geometric and arithmetic means.

2.3 The role of $quality_{it}$

When the products in scope of the price index serve common purposes, $quality_{it}$ quantifies some view on the amount of quality (as per the term's common usage) in product i at time t. But this definition is less clear for products that serve different purposes, because reasonable people can disagree about whether common usage of "quality" allows comparisons across product classes. To lessen ambiguity in such cases, $quality_{it}$ can in general be understood as quantifying some view on the amount of intrinsic value, or utility, in each product. Thus an unreliable car will typically still have a higher value of $quality_{it}$ than a perfect carrot. I settle for the label $quality_{it}$ to connect to the familiar idea of quality adjustment, and to avoid further proliferating the overlapping technical language in the literature. In particular, I copy the language of de Haan & Krsinich (2018), who use a similar concept. von Auer (2014) uses the label "transformation rate", measured in "intrinsic-worth units". Judging by a German-to-English translation in Balk (2008, p.8), Lehr (1885) favours "pleasure unit".

In equation (6), any enumeration of $quality_{it}$ is usually unique only up to a linear transformation. More precisely, if $f(\cdot)$ defines a power mean as it usually will, linear homogeneity of $\mathbb{M}^f_w[\cdot]$ implies that any enumeration of $quality_{it}$ yields the same index function as using $quality_{it}C$, where C is a strictly positive constant. Practitioners need only have views on quality relativities.

The role of $quality_{it}$ in the general model is to standardise product prices, such that within-t comparisons of quality-adjusted prices, $p_{it}/quality_{it}$, reveal differences in value for money. Hence Davies (1924) calls his equivalent of the same ratio a "dollar's worth". A prominent special case of the model arises when products all have the same quality-adjusted price within

t, in which case all ε_{it} equal zero. Equation (2) then rearranges to

$$p_{it} = quality_{it}\gamma_t \qquad (t = 1, ..., T), \tag{7}$$

where $\gamma_t = f^{-1}(\alpha_t)$. Equation (7) states the law of one price, as described in Rao & Hajargasht (2016).

And how to obtain the values of $quality_{it}$? The general model accommodates two options. The first treats the $quality_{it}$ values as revealed by an observable proxy, such as package size or some measure of relative prices. Section 3 gives examples of index functions that implicitly take this approach. The second option treats the $quality_{it}$ values as identified by extra model assumptions, as in the standard model. There, the added assumptions are in A3 ($\mathbb{E}[\boldsymbol{x_{it}}\varepsilon_{it}] = \mathbf{0}$) a special case of A3' ($\mathbb{E}_w[\varepsilon_{it}] = 0$ for all t). Those assumptions pin down the values of $quality_{it}$ because there is only ever one vector value of $\boldsymbol{\beta}$, in the term $\boldsymbol{\beta' spec_i}$, that satisfies them. With $quality_{it}$ pinned down like this, but still unobserved, practitioners can later proceed to estimating it jointly with the α_t. As discussed in Section 3, familiar estimators will be appropriate.

2.4 The role of $\boldsymbol{w_{it}}$

To borrow language from Diewert (2010), the weights w_{it} quantify the relative economic importance of product i at time t. Their role in the general model is to allow each observation to count more or less towards the target price index than others do. In formal econometric language, they help to define the population of interest that the target price index describes. I leave the mathematical details to Appendix A.1 because they are dry and taken from another econometric literature.

An equivalent implementation of the weighting could be through assumption A2', by stating that each observation is a random sample from a weighted population. Either way, the weights take effect through the term $w_{it}/\mathbb{E}[w_{it}]$, so positive linear transformations of any enumeration of w_{it} (into $w_{it}C$ for example) yield identical indexes.

Using weights has long been considered important for price index measurement (see Fisher 1922 p.43, Keynes 1930 p.78 and Griliches 1971 p.8). Heravi & Silver (2007, p.251) even write that using weights, where possible, is "axiomatic". But the general model breaks tradition because it introduces weights explicitly; the standard model has none in it. Practitioners usually introduce weights only in the estimation stage of their work, with weighted least squares (WLS). Examples of this approach arise in econometric textbooks (Berndt 1991), measurement handbooks (International Labour Office et al. 2020, International Labour Organisation et al. 2004), and countless research articles. The weights are usually functions of expenditure shares $s_{it} = p_{it}q_{it}/\sum_i p_{it}q_{it}$, where q_{it} is a transaction quantity.

The problem with this approach—and why I depart from it—is that these WLS estimators are econometrically inconsistent for the parameters defined by the standard model (Appendix A.2 gives a proof and indicative empirical example). The inconsistency arises because the expenditure shares in the weights are endogenous; they are functions of prices and thus the error terms in the price equation. This is a major departure from the types of weights that might normally be used to, say, improve estimation efficiency. Moreover, since consistency is usually considered a minimal property for good estimators, the inconsistency implies that the standard model does not justify the WLS estimators so common in the price index literature.

The normal response here might be to reject the WLS estimators and look for others that better target the parameters defined by the standard model. But the arguments for expenditure weighting, when made without reference to the standard model, are well established and sensible. So instead I question the capacities of the standard model, adding weights to make it compatible with expenditure-weighted estimation. Voltaire & Stack (1980) do something similar, but in a different model of narrow appeal. For example, it cannot handle changes in available product types.[4]

Because the arguments here are unusal, they deserve repeating at a higher level. The reasons we write down models using the likes of A1' to A3' are to i) declare useful assumptions about the data-generating process under study and ii) define which features of that process we aim to discover, in our case the parameters α_t. Without doing so we cannot sensibly proceed to estimation because we lack the means to identify good estimators. Above all, the concept of consistency is undefined without exact descriptions of the parameters interest. I show that mainstream WLS estimators are inconsistent for the parameters defined by the standard model, but only because those are rarely the true parameters of interest. Extending the model to incorporate weights, and thereby allowing other definitions for the parameters of interest, will in Section 3 reveal the mainstream estimators to be excellent.

Including weights in the general model like this might seem pedantic because it merely reinforces today's favoured approaches to estimation. It is useful because when scholars have suspected the inconsistency they have come to doubt the relevance of econometric perspectives to economic measurement, examples being Diewert (2005a) and Diewert (2010). Others have reverted to unweighted estimation, as in Feenstra (1995).[5] None of these doubts would arise if we correctly specified the parameters of interest, as the general model does.

2.5 The properties of ε_{it}

The assumptions of the general model, A1' to A3', are jointly so reasonable as to barely deserve the assumption label. This feature arises because equation (2) takes the form of

[4]Another paper with ideas overlapping with those here comes from Machado & Santos Silva (2006), notwithstanding their emphasis on quantity weights (rather than expenditure weights). They write that if the parameters of interest are defined by a model of prices for individual transactions, rather than individual products as in this paper, WLS with quantity weights is needed for econometric consistency.

[5]Likewise, in de Haan (2004) the suspected inconsistency fuels doubt about the merits of expenditure-weighted estimation. Persons (1928) first showed this line of thinking, outside of a stochastic framework.

a "saturated dummy-variable regression", meaning it contains no restrictions to functional form. Thus there always exist unique values of α_t that satisfy the assumptions (Hansen 2021 ch.2 explains such models). As per equation (3), those α_t values are quasilinear means of quality-adjusted prices in each t.

What about when adding assumptions to define $quality_{it}$, as described in Section 2.3? Are those assumptions reasonable too? Again yes, for similar reasons, exemplified in the special case that is the standard model. The relevant feature of that model is that its assumptions describe a linear "projection". So there always exist unique values of α_t and parameters $\boldsymbol{\beta}$ that satisfy A1 to A3 (Hansen 2021 ch.2 discusses projection models). Measurement practitioners often replace A3 with the more stringent special case of $\mathbb{E}[\varepsilon_{it}|\boldsymbol{x_{it}}] = 0$, in which case the standard model graduates from describing a projection to a conditional mean. But theoretical work on the equilibria of differentiated product markets, from Rosen (1974), Berry et al. (1995), and Pakes (2003), shows that the stringency is unreasonable. The stringency is also unnecessary, because the projection form still generates transparent, data-driven definitions for $quality_{it}$. In any case, nothing about the general model prevents practitioners from working with the more stringent assumption versions if they prefer to.

Note that the general model could instead have been specified using a multiplicative error representation, replacing A1' and A3' with A1" and A3":

A1"

$$f\left(\frac{p_{it}}{quality_{it}}\right) = \alpha_t \epsilon_{it} \qquad (t = 1, ..., T), \tag{8}$$

A3" For all t, the errors satisfy $\mathbb{E}_w[\epsilon_{it}] = 1$.

This representation is observationally equivalent to the additive one, because A1" and A3" yeild the same α_t as in equation (3). The error ϵ_{it} in the multiplicative representation thus maps to the error ε_{it} in the additive one according to $\varepsilon_{it} = \alpha_t \epsilon_{it} - \alpha_t$.[6] The advantages of the additive representation are pedagogical; it shortens the leap to the general model from the standard one, and makes the route to sensible estimation more obvious.

3 The general model justifies all index functions

3.1 Bilateral index functions as consistent estimators

The target index in equation (6) now has an intuitive interpretation: it measures growth in some quasilinear mean of quality-adjusted prices. Appendix A.3 shows that the sample form of this target index produces estimators that are consistent for it. The same appendix also shows that those consistent estimators equate to WLS regressions of $p_{it}/quality_{it}$ on time dummies (noting that sometimes those $quality_{it}$ values will be estimated simultaneously). In other words, consistent estimators for the target index have the form

$$\widehat{P}_{m,n}^{general} \equiv \frac{f^{-1}(\widehat{\alpha}_n^{WLS})}{f^{-1}(\widehat{\alpha}_m^{WLS})} \tag{9}$$

$$= \frac{f^{-1}\left(\sum_i \frac{w_{in}}{\sum_i w_{in}} f\left(\frac{p_{in}}{quality_{in}}\right)\right)}{f^{-1}\left(\sum_i \frac{w_{im}}{\sum_i w_{im}} f\left(\frac{p_{im}}{quality_{im}}\right)\right)}. \tag{10}$$

It turns out—and this is first central result of the paper—that the right side of equation (10) describes practically all of the bilateral price index functions in the measurement literature. (Most if not all of them are recorded in or referenced by Fisher 1922, Sato 1974, Banerjee 1983, Bryan & Cecchetti 1994, Balk 2008, von Auer 2014, and Redding & Weinstein 2020.) More precisely, the right side of (10) describes at least all of the recorded bilateral price

[6] Hence $Var(\varepsilon_{it}) = \alpha_t^2 Var(\epsilon_{it})$ for all t. So if ϵ_{it} is homoskedastic across t, as we would expect if $f(\cdot) = (\cdot)$, then ε_{it} will be heteroskedastic. Likewise, if ε_{it} is homoskedastic across t, as we would expect if $f(\cdot) = ln(\cdot)$, then ϵ_{it} will be heteroskedastic. These views on heteroskedasticity only become relevant when choosing to add identifying assumptions to pin down $quality_{it}$, i.e. when the model becomes a projection. A forthcoming discussion on multilateral indexes contains examples, especially Online Appendix A.5.

index functions that:

- treat t as discrete. This excludes a continuous time index from Divisia (1926).
- are explicit. This excludes types that are defined uniquely as the residual of a quantity index function. A prominent example is the implicit Törnqvist price index function, discussed in Diewert (1992).
- are not the esoteric bilateral types proposed by Montgomery (1937), Stuvel (1957), and Banerjee (1983).
- are not the quadratic-mean-of-order r functions from Diewert (1976), except for the special case functions from Törnqvist (1936) and Fisher (1922, p.142 no.153; also called the "ideal" index), both of which equation (10) does describe.

For a worked example, let $f(\cdot) = ln(\cdot)$, $quality_{it} = z_i \in \mathbb{R}_{++}$ (i.e. any real number that is fixed over t for each product), and $w_{it} = 0.5(s_{im} + s_{in}) \equiv w_i^{Tornqvist}$. In that case, the right side of equation (10) becomes the Törnqvist index:

$$\widehat{P}_{m,n}^{Tornqvist} = \frac{exp\left(\sum_i \frac{0.5(s_{im}+s_{in})}{\sum_i 0.5(s_{im}+s_{in})} ln\left(\frac{p_{in}}{z_i}\right)\right)}{exp\left(\sum_i \frac{0.5(s_{im}+s_{in})}{\sum_i 0.5(s_{im}+s_{in})} ln\left(\frac{p_{im}}{z_i}\right)\right)} \tag{11}$$

$$= \prod_i \left(\frac{p_{in}}{z_i}\right)^{w_i^{Tornqvist}} \prod_i \left(\frac{p_{im}}{z_i}\right)^{-w_i^{Tornqvist}} \tag{12}$$

$$= \prod_i \left(\frac{p_{in}}{p_{im}}\right)^{w_i^{Tornqvist}} . \tag{13}$$

Table 1 lists other complying bilateral functions and their settings for $f(\cdot)$, $quality_{it}$, and w_{it}. It includes types that statistical agencies use most often, based on my judgement and the results of a survey in Stoevska (2008). It also lists some for their unusual forms. For presentational purposes, it leaves out types that are quasilinear means of other indexes, such as the ideal function from Fisher (1922).[7]

Notice that many indexes match to several distinct combinations of $f(\cdot)$, $quality_{it}$, and w_{it}. To definitively identify all of the combinations for each index type is a difficult problem, not solved here. The results could be surprising. For example, Appendix A.4 includes a derivation from Bert Balk (personal correspondence) that reveals an unexpected combination for the Dutot function.

Still, it is already clear that many index types cover every possible setting for $quality_{it}$ that is constant over t. That is, the indexes are "quality-robust". Otherwise the functions tend to define $quality_{it}$ through some measure of relative prices, again constant over t. Practitioners using these indexes change quality definitions exclusively through the standard practice of chaining. An exception is the function from Redding & Weinstein (2020), which gauges $quality_{it}$ using expenditure shares that vary over t. Derived using the economic approach, the function aims to measure cost of living changes under dynamic preferences.

A noteworthy special case of the generalised index function is the "Generalised Unit Value Index Family", from von Auer (2014). The Family is the group for which $f(\cdot) = (\cdot)$, $quality_{it} = z_i \in \mathbb{R}_{++}$, and $w_{it} = quality_{it}q_{it}$, and includes Paasche and Laspeyres as examples. For each Family member, the implied quantity index always measures growth in the aggregate amount of transacted quality, an appealing feature.

[7] This also includes transitive indexes of the type introduced by Ivancic et al. (2011), as well as their predecessor indexes called GEKS. For my framework to explicitly incorporate those, the generalised target index would need to be written as a "symmetric" quasilinear mean of elements each having the same form as the right side of equation (5). Incorporating indexes using medians is also possible, by changing A3' to its quantile regression equivalent, following Koenker & Bassett Jr (1978). In that case the $f(\cdot)$ transformation would be dropped too.

Table 1: Econometric interpretations of bilateral price index functions

Index name (year)	Function ($\widehat{P}_{1,2}$)	$f(\cdot)$	$quality_{it}$	w_{it}
Dutot (1738)	$\frac{\sum_i p_{i2}}{\sum_i p_{i1}}$	$(\cdot)$	$z_i \in \mathbb{R}_{++}$	$quality_{it}$
Carli (1764)	$\frac{1}{N}\sum_i \frac{p_{i2}}{p_{i1}}$	$(\cdot)$	p_{i1}	1
$\vdots$	$\vdots$	$(\cdot)^{-1}$	p_{i2}	1
Jevons (1863)	$\prod_i \left(\frac{p_{i2}}{p_{i1}}\right)^{\frac{1}{N}}$	$ln(\cdot)$	$z_i \in \mathbb{R}_{++}$	1
Laspeyres (1871)	$\frac{\sum_i p_{i2}q_{i1}}{\sum_i p_{i1}q_{i1}}$	$(\cdot)$	$z_i \in \mathbb{R}_{++}$	$quality_{it}q_{i1}$
$\vdots$	$\vdots$	$(\cdot)$	p_{i2}	$quality_{it}q_{it}$
$\vdots$	$\vdots$	$(\cdot)^{-1}$	p_{i1}	$p_{i2}q_{i1}$
$\vdots$	$\vdots$	$(\cdot)^{-1}$	p_{i2}	$p_{i1}q_{i1}$
Paasche (1874)	$\frac{\sum_i p_{i2}q_{i2}}{\sum_i p_{i1}q_{i2}}$	$(\cdot)$	$z_i \in \mathbb{R}_{++}$	$quality_{it}q_{i2}$
$\vdots$	$\vdots$	$(\cdot)$	p_{i1}	$quality_{it}q_{it}$
$\vdots$	$\vdots$	$(\cdot)^{-1}$	p_{i2}	$p_{i1}q_{i2}$
$\vdots$	$\vdots$	$(\cdot)^{-1}$	p_{i1}	$p_{i2}q_{i2}$
Törnqvist (1936)	$\prod_i \left(\frac{p_{i2}}{p_{i1}}\right)^{0.5(s_{i1}+s_{i2})}$	$ln(\cdot)$	$z_i \in \mathbb{R}_{++}$	$0.5(s_{i1}+s_{i2})$
Lloyd (1975)-Moulton (1996)	$\left(\sum_i s_{i1}\left(\frac{p_{i2}}{p_{i1}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$	$(\cdot)^{1-\sigma}$	p_{i1}	$quality_{it}q_{i1}$
Sato (1976)-Vartia (1976)	$\prod_i \left(\frac{p_{i2}}{p_{i1}}\right)^{w_i^{SatoVartia}}$	$ln(\cdot)$	$z_i \in \mathbb{R}_{++}$	$\frac{s_{i1}-s_{i2}}{ln(s_{i1})-ln(s_{i2})}$
Redding & Weinstein (2020)	$\prod_i \left(\frac{p_{i2}}{p_{i1}}\left(\frac{s_{i2}}{s_{i1}}\right)^{\frac{1}{\sigma-1}}\right)^{\frac{1}{N}}$	$ln(\cdot)$	$s_{it}^{\frac{1}{1-\sigma}}$	1
$\vdots$	$\vdots$	$ln(\cdot)$	ψ_{it}	$\frac{s_{i1}-s_{i2}}{ln(s_{i1})-ln(s_{i2})}$

Notes: The columns labelled $f(\cdot)$, $quality_{it}$, and w_{it} give settings that, when substituted into equation (10) of the main text, give the index function in the column labelled $\widehat{P}_{1,2}$. The terms p_{it}, q_{it}, and s_{it} are prices, quantities, and within-t expenditure shares for each time-product pair. N is the sample size of products. $quality_{it} = z_i \in \mathbb{R}_{++}$ indicates that any strictly positive definitions of $quality_{it}$ that are constant overt t are admissible. σ is a consumer elasticity of substitution. The index from Redding & Weinstein (2020) is what the authors call the "common varieties" index (p.512); ψ_{it} is a time-varying taste parameter, explained further in their paper. The Dutot, Carli, Laspeyres, Paasche and Moulton attributions have all been taken on authority of Balk (2008).

3.2 Multilateral index functions as consistent estimators

The second central contribution of this paper is that equation (10) also describes practically all multilateral price index functions in the literature. (Most of them are recorded in or referenced by Hill 1997, Balk 2008, Rao & Hajargasht 2016, and Hajargasht & Rao 2019.) Possible exceptions are early types that were excluded from a taxonomy in Hill (1997).[8]

Table 2 lists some of the function types, several of which support official purchasing power parity statistics from the World Bank. The table does the measures a disservice though, hiding ingenuity behind $quality_{it}$ definitions that I have abbreviated to $\bar{p}_i$, $\widehat{p_i}$, and $\widetilde{p_i}$. Those definitions all correspond to sensible econometric estimates of each $quality_{it}$. Online Appendix A.5 details this technical result, drawing heavily on Rao & Hajargasht (2016). The main difference is that their econometric models omit the necessary weights, meaning that econometric inconsistency challenges their justification. I also give the first econometric interpretation of the $quality_{it}$ definitions implicit in the index of Geary (1958) and Khamis (1972).[9]

4 These results reveal problems and opportunities

4.1 The stochastic approach for choosing index functions

These results reveal that price index functions consistently estimate growth in some quasi-linear mean of quality-adjusted prices. The different functions are distinguished by settings for $f(\cdot)$, $quality_{it}$, and w_{it}. But which settings are ideal? Finding out would reveal the best index functions to use, yet doing so is challenging.

Consider, for example, the task of discriminating between bilateral functions on the basis of

[8] References for these possible exceptions are in Balk (2008, p.35), starting with Theil (1960) and Kloek & De Wit (1961).

[9] Rao & Selvanathan (1992) offer an econometric interpretation but treat the price index as known.

Table 2: Econometric interpretations of multilateral functions

Index name (year, type)	Function ($\widehat{P}_{m,n}$)	$f(\cdot)$	$quality_{it}$	w_{it}
Walsh (1901)	$\prod_i \left(\frac{p_{in}}{p_{im}}\right)^{\frac{1}{T}\sum_t s_{it}}$	$ln(\cdot)$	$z_i \in \mathbb{R}_{++}$	$\frac{1}{T}\sum_t s_{it}$
Van Ijzeren (1956)	$\frac{\sum_i p_{in}\bar{q}_i}{\sum_i p_{im}\bar{q}_i}$	$(\cdot)$	$z_i \in \mathbb{R}_{++}$	$quality_{it}\bar{q}_i$
$\vdots$	$\vdots$	$(\cdot)^{-1}$	p_{im}	$p_{in}\bar{q}_i$
Geary (1958)-Khamis (1972)	$\frac{\sum_i p_{in}q_{in}}{\sum_i \bar{p}_i q_{in}} \Big/ \frac{\sum_i p_{im}q_{im}}{\sum_i \bar{p}_i q_{im}}$	$(\cdot)$	$\bar{p}_i$	$quality_{it}q_{it}$
$\vdots$	$\vdots$	$(\cdot)^{-1}$	$\bar{p}_i$	$p_{it}q_{it}$
Rao (1990)	$\prod_i \left(\frac{p_{in}}{\widehat{p}_i}\right)^{s_{in}} \prod_i \left(\frac{p_{im}}{\widehat{p}_i}\right)^{-s_{im}}$	$ln(\cdot)$	$\widehat{p}_i$	s_{it}
Hajargasht & Rao (2010, 1)	$\sum_i s_{in}\frac{p_{in}}{\widetilde{p}_i} \Big/ \sum_i s_{im}\frac{p_{im}}{\widetilde{p}_i}$	$(\cdot)$	$\widetilde{p}_i$	s_{it}

Notes: The columns labelled $f(\cdot)$, $quality_{it}$, and w_{it} give settings that, when substituted into equation (10) of the main text, give the index function in column labelled $\widehat{P}_{m,n}$. The terms p_{it}, q_{it}, and s_{it} are prices, quantities, and within-t expenditure shares for each time-product pair. $quality_{it} = z_i \in \mathbb{R}_{++}$ indicates that any strictly positive definitions of $quality_{it}$ that are fixed across t are admissible. Precise definitions of $\bar{p}_i$, $\widehat{p}_i$, and $\widetilde{p}_i$ are available in Appendix A.5. See Hill (1997) for details on $\bar{q}_i$. The Van Ijzeren attribution is taken on authority of Balk (2008).

$quality_{it}$. Table 1 shows most of the functions to be quality-robust under certain settings for $f(\cdot)$ and w_{it}. Taking this feature to be ideal, we cannot use $quality_{it}$ as a unique basis on which to discriminate between those functions. Moreover, other settings for $quality_{it}$, using relative prices, also make economic sense by appealing to the principle of revealed preference. No obvious deal-breakers arise in index settings for $quality_{it}$.

What about discriminating on the basis of $f(\cdot)$? We cannot use goodness-of-fit criteria, because equation (2) is saturated with dummy variables and the different quasilinear means already minimise their respective quasilinear loss functions (de Carvalho 2016). For general econometric use, Gorajek (2019) offers other selection criteria for quasilinear means, focussing on the needs of relevant policymakers. But index functions simultaneously serve many different policymakers, with diverse needs. Here too there are no obvious deal-breakers.

One might try backing out ideal settings for $f(\cdot)$, $quality_{it}$, and w_{it} using the index functions recommended by the economic or axiomatic approaches. But doing so would presuppose the functions one wants to discover; it would not reveal anything new.

The bottom line here is that unless scholars can find ideal settings for $f(\cdot)$, $quality_{it}$, and w_{it} without relying on the economic or axiomatic approaches, the stochastic approach to choosing index functions does not actually help choose index functions. Left to its own, and in a more complete form such as the general model, the stochastic approach can justify all common functions. Judging by International Labour Office et al. (2020), statistical agencies already rely little on stochastic approach to justify index function choices. The general model suggests it would be productive to disregard it for that purpose altogether.

4.2 The stochastic approach for measuring index uncertainty

As explained by Diewert (2010) and Manski (2015), for over a century scholars have argued for statistical agencies to be more transparent about the uncertainty in macroeconomic aggregates, potentially through the use of confidence intervals. In turn, scholars of the stochastic approach have argued it to be the best available framework for calculating such intervals. Wherever index numbers can legitimately be understood as textbook econometric estimators, practitioners need only use textbook econometric tools. For example, Rao & Hajargasht (2016) show how to calculate the intervals using asymptotic standard errors. Rao & Selvanathan (1992) use the bootstrap. Other approaches appear in Clements & Izan (1981; 1987).

By showing that in fact all common price indexes can be understood as textbook econometric estimators, a contribution of the general model is to reveal that one always has the means to calculate sensible confidence intervals.

Of the econometric methods available, the bootstrap is already understood to require the most computational power and provide the most accuracy. It also offers an important advantage for statistical agencies: without much difficulty, it can produce sensible intervals for complex functions of estimators. This helps because statistical agencies routinely combine indexes, such as when chaining, or when aggregating the elementary indexes of a consumer price index. In all cases, the same resampling logic applies. I do not give more detail because many textbooks do already.[10]

4.3 The Time-dummy Hedonic Method of Quality Adjustment

According to Triplett (2004), adjusting for quality change that comes from market entry and exit of product types has "long been recognised as perhaps the most serious problem in estimating price indexes" (p.11). Moulton (2018) explains how successive investigations have estimated the problem to account for the largest source of measurement bias in the US consumer price index. And Aghion et al. (2019) argue that those investigations do not even capture the full extent of the problem.

An old solution, but possible only with data on product specifications, is time-dummy hedonic regression. As discussed already, the target indexes then equal $exp(\alpha_n - \alpha_m)$ from the special case of the standard model in which vector $\boldsymbol{spec_i}$ contains product attributes. The implied $quality_{it}$ definitions are the unique $\boldsymbol{\beta' spec_i}$ that satisfy A3. Least squares methods simultaneously yield estimates of $quality_{it}$ and the price indexes, which, in turn, are special cases of the generalised index in equation (10). Thus the time-dummy hedonic method has the same conceptual foundations as index functions, especially when care is taken to properly model weights.

But experts criticise the time-dummy hedonic solution and have stifled broader take-up at

[10]Hansen (2021) is an excellent example. Measurement practitioners will need the "panel bootstrap", to avoid creating product mismatches with the resampling algorithm.

statistical agencies. For example, answering a request by the US Bureau of Labor Statistics (BLS), a panel of experts writes:

> "Recommendation 4-4: BLS should not allocate resources to the [time-dummy hedonic method] (unless work on other hedonic methods generates empirical evidence that characteristic parameter stability exists for some products)." (National Research Council 2002, p.143)

Similarly:

> "The main concern with the use of the hedonic time dummy index approach ... is that by construction, it constrains the parameters on the characteristic variables to be the same." (Diewert et al. 2009, p.188)

The criticism is that restricting the coefficients on product specifications to be unchanged over time, even just two adjacent periods, is unreasonable. Or, to use the language of the general model, the criticism is that restricting the definition of $quality_{it}$ to be unchanged over time contradicts ever-changing views about product quality. Several papers reject coefficient stability over adjacent periods for the computer market in the U.S. and echo the criticism (Berndt & Rappaport 2001, Pakes 2003, Reis & Santos Silva 2006).

I argue that the criticism is flawed, for two reasons:

1. The results in Tables 1 and 2 show that index functions also use time-constant measures of $quality_{it}$. Practitioners using these indexes incorporate changing views on quality exclusively through the practice of chaining, a practice that also works with the time-dummy method. So unless the criticism about time-constant $quality_{it}$ is intended to cover index functions as well, it is not a logically consistent basis on which to challenge the time-dummy hedonic method. Putting this another way, restricting the $quality_{it}$ measures to be unchanged over time makes the time-dummy hedonic method (for when

product types are dynamic) cohere with index functions (for when product types are static).

2. For that matter, index functions restrict $quality_{it}$ in this way for good reason. Without the restriction, measured price changes could stem entirely from updated views about quality, and functions would fail critical tests in the axiomatic approach (see International Labour Office et al. 2020, p.180, especially the proportionality test). Likewise, the economic approach has a long tradition of holding preferences fixed over comparison periods, starting with Konüs (1939, translated from a 1924 version). In a challenge of the dynamic-preferences index from Redding & Weinstein (2020), Kurtzon (2020) highlights problems that arise when departing from this norm.

These arguments matter because alternatives to the time-dummy hedonic method—most notably hedonic imputation—often yield different economic narratives. For example, on a sample of British desktop computers, Diewert et al. (2009) show that the difference between the imputation and time-dummy methods accumulates to between 3.2 and 7.5 percentage points over a year. The push to use these alternatives is impactful and poorly justified.

4.4 Other Applications in Price Index Measurement

I have used the general model for two other methodological investigations. But since the results reinforce the merits of existing practices, I leave the details to online appendices. Summaries of each:

- This paper has so far followed the modelling convention in economic measurement by assuming that each product has a single price in each time period t. Usually the assumption is unrealistic, because t is an area, not a point. Price changes can fall within its boundaries. So how can index functions handle breaches of the single price assumption? I use the general model to tackle this question in Online Appendix

A.6. The findings support the prevailing solution of collapsing multiple prices into "unit values", even though that solution was initially proposed with computational convenience as a central motivation.[11]

- The general model reveals sensible new functional form possibilities for time-dummy hedonic regression, mostly through different $f(\cdot)$. The geometric mean ($f(\cdot) = ln(\cdot)$) is standard, but not necessarily best, especially for statistical agencies seeking a coherence with index functions that use other means. In discussing the general model with others I have been asked how to estimate hedonic indexes using these non-standard choices of $f(\cdot)$. I estimate one in Online Appendix A.7, and include the replication code in the supporting material. My index estimates look a lot like those coming from the equivalent geometric mean index, but are challenging to calculate objectively because computational methods are required. Different starting values of the solution algorithms can cause large changes to the estimates. I conclude that departing from the geometric mean will often be impractical for the time-dummy hedonic method.

5 Conclusion

It turns out that price index functions share a common interpretation; practically all of them consistently estimate a change in some quasilinear mean of quality-adjusted prices. The different options are distinguished by a choice of quasilinear mean, a definition of quality, and a stance on what constitutes the population of interest (weighting).

This new interpretation reveals problems and opportunities in economic measurement. First, the stochastic approach to choosing index functions is unhelpful for choosing index functions, because in its complete form it can justify all of them. Second, the same feature of the

[11] To clarify, I investigate the use of unit values as opposed to "unit value indexes", which are simple ratios of two unit values. My language follows International Labour Office et al. (2020).

stochastic approach makes it possible to calculate confidence intervals for any index type. Third, the literature uses flawed arguments for swapping the time-dummy hedonic method of quality adjustment with hedonic imputation.

A Technical Appendices

A.1 Weights and the population of interest

By definition,

$$\begin{aligned} \mathbb{E}_w[\varepsilon_{it}] &\equiv \mathbb{E}[\varepsilon_{it} w_{it}] \frac{1}{\mathbb{E}[w_{it}]} &(14)\\ &\equiv \int\int \varepsilon_{it} w_{it} g(\varepsilon_{it}, w_{it})\ \mathrm{d}\varepsilon_{it} \mathrm{d}w_{it}\ \frac{1}{\mathbb{E}[w_{it}]} &(15)\end{aligned}$$

where $g(\varepsilon_{it}, w_{it})$ is a density function. Now copying the logic of the method of weighted importance sampling in Kroese et al. (2011, p.368), we can define a new density function $h(\varepsilon_{it}, w_{it}) \equiv w_{it} g(\varepsilon_{it}, w_{it}) \mathbb{E}[w_{it}]^{-1}$ and rewrite equation (15) as

$$\mathbb{E}_w[\varepsilon_{it}] \equiv \int\int \varepsilon_{it} h(\varepsilon_{it}, w_{it})\ \mathrm{d}\varepsilon_{it} \mathrm{d}w_{it}. \tag{16}$$

The same logic carries over to the α_t parameters, which equation (3) equates to the value $\mathbb{E}_w[f(p_{it}/quality_{it})]$. Thus the weighting has changed the stated population of interest, via the density functions underlying $\mathbb{E}_w[\cdot]$.

A.2 Inconsistent textbook estimation

Let $\boldsymbol{\delta}$ and $\boldsymbol{x_{it}}$ be vector shorthand for the full set of coefficients and regressors that are implicit in the standard model. N is the sample size. Then

$$\widehat{\boldsymbol{\delta}}^{\boldsymbol{WLS}} = \left(\sum_{it} (\boldsymbol{x_{it}} w_{it} \boldsymbol{x'_{it}}) \right)^{-1} \sum_{it} (\boldsymbol{x_{it}} w_{it} ln(p_{it})) \tag{17}$$

$$= \left(\sum_{it} \frac{1}{N} (\boldsymbol{x_{it}} w_{it} \boldsymbol{x'_{it}}) \right)^{-1} \sum_{it} \frac{1}{N} (\boldsymbol{x_{it}} w_{it} ln(p_{it})). \tag{18}$$

Now applying the Law of Large Numbers and the Continuous Mapping Theorem,

$$plim_{N\to\infty}\, \widehat{\boldsymbol{\delta}}^{\boldsymbol{WLS}} = (\mathbb{E}[\boldsymbol{x_{it}} w_{it} \boldsymbol{x'_{it}}])^{-1} \mathbb{E}[\boldsymbol{x_{it}} w_{it} ln(p_{it})] \tag{19}$$

$$\Rightarrow plim_{N\to\infty}\, \widehat{\boldsymbol{\delta}}^{\boldsymbol{WLS}} - \boldsymbol{\delta} = (\mathbb{E}[\boldsymbol{x_{it}} w_{it} \boldsymbol{x'_{it}}])^{-1} \mathbb{E}[\boldsymbol{x_{it}} w_{it} \varepsilon_{it}] \tag{20}$$

$$= (\mathbb{E}[\boldsymbol{x_{it}} w_{it} \boldsymbol{x'_{it}}])^{-1} \mathbb{C}ov[\boldsymbol{x_{it}} \varepsilon_{it}, w_{it}]. \tag{21}$$

The right side of equation (20) is a weighted linear projection of the errors on the regressors. Since expenditure weights are functions of the errors (through prices), the second expectation term does not in general equal zero.

Strictly speaking, one cannot demonstrate the inconsistency empirically because doing so would require infinite sample sizes. But studies with enormous cross sections do offer indicative evidence. Notable here is work by Fox & Syed (2016), which contains scanner data comparisons of indexes from Jevons (1863) and Törnqvist (1936). Though not the intention of Fox & Syed, their comparisons speak to the inconsistency because the literature proposes both indexes as estimators for versions of the standard model in which $T = 2$ and product types are static, the only difference being that Törnqvist uses (endogenous) expenditure weights. Drawing on over 20 million observations on basic household products, sold across six major US cities, Fox & Syed (2016) show that the gap between the Jevons and Törnvqvist indexes accumulates to 12 percentage points over 11 years.

A.3 WLS consistency for the new target index

$$\begin{bmatrix} \widehat{\alpha}_1^{WLS} \\ \vdots \\ \widehat{\alpha}_T^{WLS} \end{bmatrix} = \begin{bmatrix} \sum_i w_{i1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sum_i w_{iT} \end{bmatrix}^{-1} \begin{bmatrix} \sum_i w_{i1}\, f\left(\frac{p_{i1}}{quality_{i1}}\right) \\ \vdots \\ \sum_i w_{iT}\, f\left(\frac{p_{iT}}{quality_{iT}}\right) \end{bmatrix} \tag{22}$$

$$\iff \qquad \widehat{\alpha}_t^{WLS} = \sum_i \frac{w_{it}}{\sum_i w_{it}} f\left(\frac{p_{it}}{quality_{it}}\right) \qquad \text{for all } t \tag{23}$$

$$= \left(\sum_i \frac{w_{it}}{N}\, f\left(\frac{p_{it}}{quality_{it}}\right)\right) \frac{N}{\sum_i w_{it}} \qquad \text{for all } t \tag{24}$$

Now applying the Law of Large Numbers and the Continuous Mapping Theorem,

$$plim_{N\to\infty}\, \widehat{\alpha}_t^{WLS} = \mathbb{E}\left[f\left(\frac{p_{it}}{quality_{it}}\right) w_{it}\right] \mathbb{E}\left[w_{it}\right]^{-1} \qquad \text{for all } t \tag{25}$$

$$\equiv \mathbb{E}_w\left[f\left(\frac{p_{it}}{quality_{it}}\right)\right] \qquad \text{for all } t \tag{26}$$

$$= \alpha_t \qquad \text{for all } t \tag{27}$$

$$\iff \qquad plim_{N\to\infty} \frac{\widehat{\alpha}_n^{WLS}}{\widehat{\alpha}_m^{WLS}} = \frac{\alpha_n}{\alpha_m} \qquad \text{for all } m, n, \tag{28}$$

where the step from equation (26) to (27) uses assumption A3', from the description of the general model in the body text.

A.4 An unexpected Dutot index interpretation

Aside from the combination listed in Table 1, the Dutot price index is defined by the combination $f(\cdot) = ln(\cdot)$, $quality_{it} = z_i \in \mathbb{R}_{++}$, and $w_{it} = l(p_{i2}/\boldsymbol{p_2}, p_{i1}/\boldsymbol{p_1})$, where $l(\cdot,\cdot)$ is the logarithmic mean, defined as $l(a,b) = (b-a)/(ln(b) - ln(a))$, and $\boldsymbol{p_t} = \sum_i p_{it}/N$.

To explain, I start with the identity

$$\sum_i \left(\frac{p_{i2}}{\sum_i p_{i2}} - \frac{p_{i1}}{\sum_i p_{i1}} \right) = 0 \tag{29}$$

$$\iff \quad \sum_i l\left(\frac{p_{i2}}{\sum_i p_{i2}}, \frac{p_{i1}}{\sum_i p_{i1}} \right) \left(ln\left(\frac{p_{i2}}{p_{i1}} \right) - ln\left(\frac{\sum_i p_{i2}}{\sum_i p_{i1}} \right) \right) = 0 \tag{30}$$

$$\iff \quad ln\left(\frac{\sum_i p_{i2}}{\sum_i p_{i1}} \right) = \sum_i \left(\frac{l\left(\frac{p_{i2}}{\sum_i p_{i2}}, \frac{p_{i1}}{\sum_i p_{i1}} \right)}{\sum_i l\left(\frac{p_{i2}}{\sum_i p_{i2}}, \frac{p_{i1}}{\sum_i p_{i1}} \right)} \right) ln\left(\frac{p_{i2}}{p_{i1}} \right). \tag{31}$$

Applying linear homogeneity of the logarithmic mean,

$$ln\left(\frac{\sum_i p_{i2}}{\sum_i p_{i1}} \right) = \sum_i \left(\frac{l\left(\frac{p_{i2}}{\boldsymbol{p_2}}, \frac{p_{i1}}{\boldsymbol{p_1}} \right)}{\sum_i l\left(\frac{p_{i2}}{\boldsymbol{p_2}}, \frac{p_{i1}}{\boldsymbol{p_1}} \right)} \right) ln\left(\frac{p_{i2}}{p_{i1}} \right) \tag{32}$$

$$\iff \quad \frac{\sum_i p_{i2}}{\sum_i p_{iv}} = \prod_i \left(\frac{p_{i2}}{p_{i1}} \right) \wedge \left(\frac{l\left(\frac{p_{i2}}{\boldsymbol{p_2}}, \frac{p_{i1}}{\boldsymbol{p_1}} \right)}{\sum_i l\left(\frac{p_{i2}}{\boldsymbol{p_2}}, \frac{p_{i1}}{\boldsymbol{p_1}} \right)} \right). \tag{33}$$

Now using the same logic as in equations (11) to (13) (in reverse), the rest is clear.

A.5 Multilateral index functions as moment estimators (ONLINE ONLY)

Many multilateral indexes treat $quality_{it}$ as a product fixed effect to be estimated. Assumption A3' then needs to change; otherwise we are imprecise about what exactly the fixed effects are that we wish to estimate. So here I replace A3' with a pair of assumptions, A3"' and A4"'. They are stronger than necessary, but simple and use the same overall logic as Rao & Hajargasht (2016). Later I will show a relaxed version that justifies the same indexes.

The full set of new assumptions is

A1"'

$$f\left(\frac{p_{it}}{\lambda_i}\right) = \alpha_t + \varepsilon_{it} \qquad (t = 1, ..., T), \tag{34}$$

where λ_i is a new product fixed effect.

A2"' Each $(p_{it}, \boldsymbol{x_{it}})$ is a random sample from the time-product population, where vector $\boldsymbol{x_{it}}$ contains all of the implied regressors in equation (34) (the product and time dummies).

A3"' The regressors are strictly exogenous, that is $E_w[\varepsilon_{it}|\boldsymbol{x_{it}}] = 0$.

A4"' There is conditional heteroskedasticity of the form

$$Var_w(\varepsilon_{it}|\boldsymbol{x_{it}}) \equiv \mathbb{E}_w\left[(\varepsilon_{it} - \mathbb{E}_w[\varepsilon_{it}|\boldsymbol{x_{it}}])^2|\boldsymbol{x_{it}}\right] \tag{35}$$

$$\propto h_{it}. \tag{36}$$

Now let $f(x) = x$, $w_{it} = q_{it}\lambda_i$ and $h_{it} = \alpha_t^2$. (Since the model is now in raw levels, that is $f(x) = x$, the heteroskedasticity assumption is natural.) A3"' then implies that

$$\mathbb{E}_w\left[\frac{p_{it}}{\lambda_i} - \alpha_t \middle| \boldsymbol{x_{it}}\right] \equiv \mathbb{E}_w\left[r(p_{it}, \boldsymbol{x_{it}}, \boldsymbol{\delta})|\boldsymbol{x_{it}}\right] \tag{37}$$

$$= 0, \tag{38}$$

where $\boldsymbol{\delta}$ is vector shorthand for the set of coefficients (the λ_i and α_t) in the model.

Following Wooldridge (2010, p.542), the efficient method of moments estimators for these coefficients solve

$$\sum_{it} \frac{1}{\alpha_t^2}\, \mathbb{E}_w[\nabla_{\boldsymbol{\delta}} r(p_{it}, \boldsymbol{x_{it}}, \boldsymbol{\delta})|\boldsymbol{x_{it}}]\; r(p_{it}, \boldsymbol{x_{it}}, \hat{\boldsymbol{\delta}}) w_{it} = \mathbf{0}. \tag{39}$$

where the w_{it} is again a correction from the importance sampling literature, necessary because the sample is drawn randomly from the population of product-time pairs, not the population governing $\mathbb{E}_w[\cdot]$.

But note that

$$\mathbb{E}_w \left[\left. \frac{\partial r(p_{it}, \boldsymbol{x_{it}}, \boldsymbol{\delta})}{\partial \alpha_t} \right| \boldsymbol{x_{it}} \right] = -1, \tag{40}$$

and

$$\begin{aligned} \mathbb{E}_w \left[\left. \frac{\partial r(p_{it}, \boldsymbol{x_{it}}, \boldsymbol{\delta})}{\partial \lambda_i} \right| \boldsymbol{x_{it}} \right] &= \mathbb{E}_w \left[\left. -\frac{p_{it}}{\lambda_i^2} \right| \boldsymbol{x_{it}} \right] & (41) \\ &= -\frac{\alpha_t}{\lambda_i}, & (42) \end{aligned}$$

which contains the unknown parameters α_t and λ_i. The feasible method of moments estimators use $\widehat{\alpha}_t$ and $\widehat{\lambda}_i$ instead. The system of equations resulting from (39) is then

$$\sum_t \frac{q_{it}}{\widehat{\alpha}_t} \left(\frac{p_{it}}{\widehat{\lambda}_i} - \widehat{\alpha}_t \right) = 0 \quad \forall i \qquad \iff \qquad \widehat{\lambda}_i = \sum_t \frac{p_{it} q_{it}}{\widehat{\alpha}_t} \left(\sum_t q_{it} \right)^{-1} \quad \forall i \tag{43}$$

and

$$\sum_i q_{it} \widehat{\lambda}_i \left(\frac{p_{it}}{\widehat{\lambda}_i} - \widehat{\alpha}_t \right) = 0 \quad \forall t \qquad \iff \qquad \widehat{\alpha}_t = \frac{\sum_i p_{it} q_{it}}{\sum_i q_{it} \widehat{\lambda}_i} \quad \forall t. \tag{44}$$

This is the same as the system of Geary (1958) and Khamis (1972). The result contrasts one by Rao & Hajargasht (2016), who argue that inefficient estimation weights are needed to generate that index. In my framework, which includes weights in the baseline model, the estimation weights are justified on consistency grounds (copying the logic of Section 2.4).

Table 3 provides the settings for $f(\cdot)$, w_{it} and h_{it} needed to generate the other multilateral indexes considered in Rao & Hajargasht (2016), using the same method as above. Note also that solving these systems of equations requires a normalisation; Rao & Hajargasht (2016) just set one of the α_t to 1.

Table 3: Method of moments interpretations for multilateral functions

Index name (year, type)	$f^{-1}(\widehat{\alpha}_t)$, $\widehat{\lambda}_i$	$f(\cdot)$	w_{it}	h_{it}
Dutot-style	$\frac{\sum_i p_{it}}{\sum_i \lambda_i}$, $\frac{1}{T}\sum_t \left(\frac{p_{it}}{\alpha_t}\right)$	$(\cdot)$	λ_i	α_t^2
Harmonic	$\left(\frac{1}{N}\sum_i \left(\frac{\lambda_i}{p_{it}}\right)\right)^{-1}$, $\left(\frac{1}{T}\sum_t \left(\frac{1}{\alpha_t p_{it}}\right)\right)^{-1}$	$(\cdot)^{-1}$	1	α_t^2
Geometric	$\prod_i \left(\frac{p_{it}}{\lambda_i}\right)^{\frac{1}{N}}$, $\prod_t \left(\frac{p_{it}}{exp(\alpha_t)}\right)^{\frac{1}{T}}$	$ln(\cdot)$	1	1
Geary (1958)-Khamis (1972)	$\frac{\sum_i p_{it}q_{it}}{\sum_i \lambda_i q_{it}}$, $\frac{\sum_t p_{it}q_{it}\alpha_t^{-1}}{\sum_t q_{it}}$	$(\cdot)$	$\lambda_i q_{it}$	α_t^2
Iklé (1972)	$\left(\sum_i s_{it}\left(\frac{\lambda_i}{p_{it}}\right)\right)^{-1}$, $\left(\sum_t s_{it}\left(\frac{1}{\alpha_t p_{it}}\right)\right)^{-1}$	$(\cdot)^{-1}$	$p_{it}q_{it}$	α_t^2
Rao (1990)	$\prod_i \left(\frac{p_{it}}{\lambda_i}\right)^{s_{it}}$, $\prod_t \left(\frac{p_{it}}{exp(\alpha_t)}\right)^{s_{it}}$	$ln(\cdot)$	$p_{it}q_{it}$	1
Hajargasht & Rao (2010, 1)	$\sum_i s_{it}\left(\frac{p_{it}}{\lambda_i}\right)$, $\sum_t s_{it}\left(\frac{p_{it}}{\alpha_t}\right)$	$(\cdot)$	$p_{it}q_{it}$	α_t^2
Hajargasht & Rao (2010, 2)	$\frac{1}{N}\sum_i \left(\frac{p_{it}}{\lambda_i}\right)$, $\frac{1}{T}\sum_t \left(\frac{p_{it}}{\alpha_t}\right)$	$(\cdot)$	1	α_t^2

Notes: The columns labelled $f(\cdot)$, w_{it} and h_{it} give settings that, when substituted into the model consisting of assumptions A1"' to A4"' of this appendix, justify the method of moment estimators (and index functions) in the middle column. The terms p_{it}, q_{it} and s_{it} are prices, quantities, and within-t expenditure shares for each time-product pair. N is sample size of observed products.

Finally, even if assumptions A3"' and A4"' are wrong, the same estimators are still consistent for population projection parameters. In other words, in the Geary-Khamis case, A3"' and A4"' could be replaced with the less stringent moment conditions

$$\mathbb{E}_w\left[\frac{\varepsilon_{it}}{\alpha_t\lambda_i}\right] = 0 \quad \forall i \qquad \text{and} \qquad \mathbb{E}_w\left[\varepsilon_{it}\right] = 0 \quad \forall t, \tag{45}$$

which generate the same estimating equations.

A.6 The merits of unit values (ONLINE ONLY)

As discussed in the main text, throughout this paper I have followed the modelling convention in economic measurement and assumed that each time-product pair has a single price. Usually the assumption is unrealistic, because t is an area, not a point. Price changes can fall within its boundaries. So how can measurement methods handle breaches of the single price assumption?

Standard practice is to collapse multiple prices into a single one using "unit values" (see International Labour Office et al. 2020 and International Labour Organisation et al. 2004). The unit values equal the total measured expenditure on each observed time-product pair divided by the measured number of transactions in that pair. In other words, they are quantity-weighted, arithmetic sample means.[12] The solution has been justified for being simple computationally, needing little information, and ordinarily producing results that are imperceptibly different from other types of sample means (Fisher 1922, p.318). But with scanner data becoming more accessible, and computers now so powerful, we are able to explore more information-intensive options.

The general model helps in identifying the ideal option here because it does not rely on price relatives. In doing so it frees us from having to assume single prices for time-product pairs. To write the model in terms of individual transactions we just have one new decision to make: how should we sample and weight the individual transactions?

The solution I suggest respects the stances on economic importance taken by existing index functions. In particular, I propose weighting schemes that (i) preserve the economic importance of each time-product pair implied in existing index functions, and at the same time (ii) allocate economic importance evenly to the transactions *within* each time-product pair.

[12] When transaction numbers are unavailable, it is common to take unweighted arithmetic averages of observed prices instead.

The general model, in terms of individual transactions, is then defined by A1* to A3*.

A1*

$$f\left(\frac{p_{ijt}}{quality_{it}}\right) = \alpha_t + \varepsilon_{ijt} \qquad (t = 1, ..., T; j = 1, ..., J), \tag{46}$$

where the j subscript tracks transactions.

A2* Starting with a random sample from the time-product population, a further J random transactions for each time-product pair are sampled to obtain each $(p_{ijt}, quality_{it})$.

A3* for all t, $E_w[\varepsilon_{ijt}] = 0$, where $w_{ijt} = w_{it}$ for all j. That is, the weights are fixed across j within each product-time pair.

The target index of quality-adjusted price growth from time period $t = m$ to $t = n$ is then

$$P^{general}_{m,n} \equiv \frac{f^{-1}(\alpha_n)}{f^{-1}(\alpha_m)} \tag{47}$$

$$= \frac{f^{-1}\left(\mathbb{E}_w\left[f\left(\frac{p_{ijn}}{quality_{in}}\right)\right]\right)}{f^{-1}\left(\mathbb{E}_w\left[f\left(\frac{p_{ijm}}{quality_{im}}\right)\right]\right)}. \tag{48}$$

And the generalised estimator becomes

$$\widehat{P}^{general}_{m,n} \equiv \frac{f^{-1}(\widehat{\alpha}^{WLS}_n)}{f^{-1}(\widehat{\alpha}^{WLS}_m)} \tag{49}$$

$$= \frac{f^{-1}\left(\sum_{ij} \frac{w_{ijn}}{\sum_{ij} w_{ijn}} f\left(\frac{p_{ijn}}{quality_{in}}\right)\right)}{f^{-1}\left(\sum_{ij} \frac{w_{ijm}}{\sum_{ij} w_{ijm}} f\left(\frac{p_{ijm}}{quality_{im}}\right)\right)} \tag{50}$$

$$= \frac{f^{-1}\left(\sum_{i} \frac{w_{in}}{\sum_{i} w_{in}} \left\{\frac{1}{J}\sum_j f\left(\frac{p_{ijn}}{quality_{in}}\right)\right\}\right)}{f^{-1}\left(\sum_{i} \frac{w_{im}}{\sum_{i} w_{im}} \left\{\frac{1}{J}\sum_j f\left(\frac{p_{ijm}}{quality_{im}}\right)\right\}\right)} \tag{51}$$

By extension of equation (10) of the main text, price indexes that use the unit values solution have the form

$$\widehat{P}_{m,n}^{general} = \frac{f^{-1}\left(\sum_i \frac{w_{in}}{\sum_i w_{in}} \left\{ f\left(\frac{1}{J}\sum_j \frac{p_{ijn}}{quality_{in}}\right)\right\}\right)}{f^{-1}\left(\sum_i \frac{w_{im}}{\sum_i w_{im}} \left\{ f\left(\frac{1}{J}\sum_j \frac{p_{ijm}}{quality_{im}}\right)\right\}\right)} \tag{52}$$

The two solutions are equivalent if

$$\frac{1}{J}\sum_j \frac{p_{ijt}}{quality_{it}} = f^{-1}\left(\frac{1}{J}\sum_j f\left(\frac{p_{ijt}}{quality_{it}}\right)\right) \qquad \forall i, t. \tag{53}$$

where the left side formally defines the unit value for time-product pair (i,t). The equality holds when the index is one for which $f(\cdot) = (\cdot)$. Otherwise, the equality need not hold, resulting in an error. This result should be unsurprising; the unit value method takes a position on the appropriate measure of central tendency without considering the choice of $f(\cdot)$, which is also a position on the appropriate measure of central tendency.

I could provide simulations here for which the errors are large, but the parameters dictating the sizes of the differences are unfamiliar, making it difficult to judge the value of the exercise. Instead I provide two extreme empirical examples below, both of which still result in only trivial differences. The bottom line is that unit values still appear to be excellent, practical solutions.

The first example calculates Törnqvist-type price indexes for the top three cryptocurrencies on an exchange called Coinbase, at three different index frequencies (Figure 1). The example is extreme because the cryptocurrencies often have large swings in prices within the periods that define the index frequencies. Thus the unit values method has to do a lot of work to summarise prices of each type of cryptocurrency. Still, the differences between the unit values method and the indexes justified by my solution are almost imperceptible.[13]

[13]Note that to calculate all of the cryptocurrency indexes I start with hourly unit values, i.e. the raw data are not quite at the tick level.

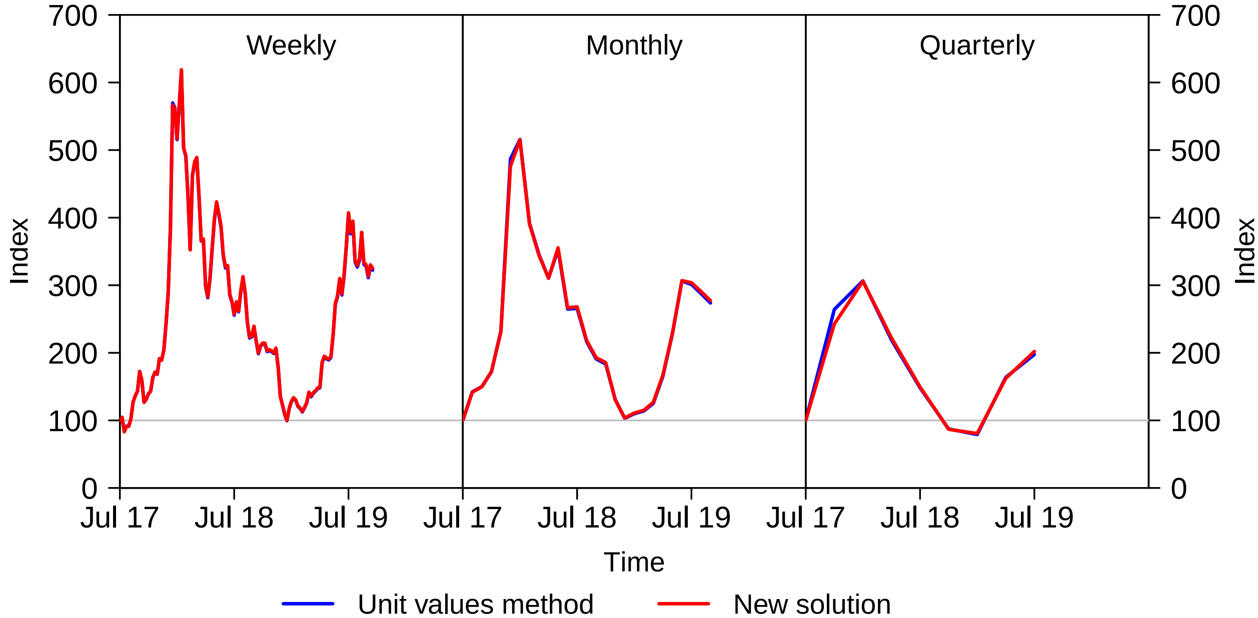

Figure 1: Törnqvist price indexes for cryptocurrencies

Notes: The cryptocurrencies in scope are Bitcoin, Ethereum, and Litecoin, all traded on the Coinbase exchange. The unit values method is standard. The new solution is described formally in the body text. Despite the extreme price volatility, the two methods give almost identical results.
Sources: Author's calculations; CryptoDataDownload.com

The second example calculates Törnqvist-type price indexes for laundry detergent products sold by a now-defunct Chicago franchise called Dominick's (Figure 2). The example is extreme for its so-called chain drift, evident in the different index trends across the weekly and monthly panels.[14] Again, the differences between the unit values method and my solution are almost imperceptible.[15]

The script to replicate these figures is in the supporting material online.

[14]de Haan & Van der Grient (2011) were first to notice this feature of laundry detergent price indexes, in a Dutch setting.

[15]To calculate all of the laundry detergent indexes I start with weekly unit values at the product-store level, i.e. the raw data are not quite at the transaction level. The raw data for week 219 are missing in the underlying dataset, so I repeat the data from week 218. I also drop products without matching observations in comparison periods, to calculate a matched-product index.

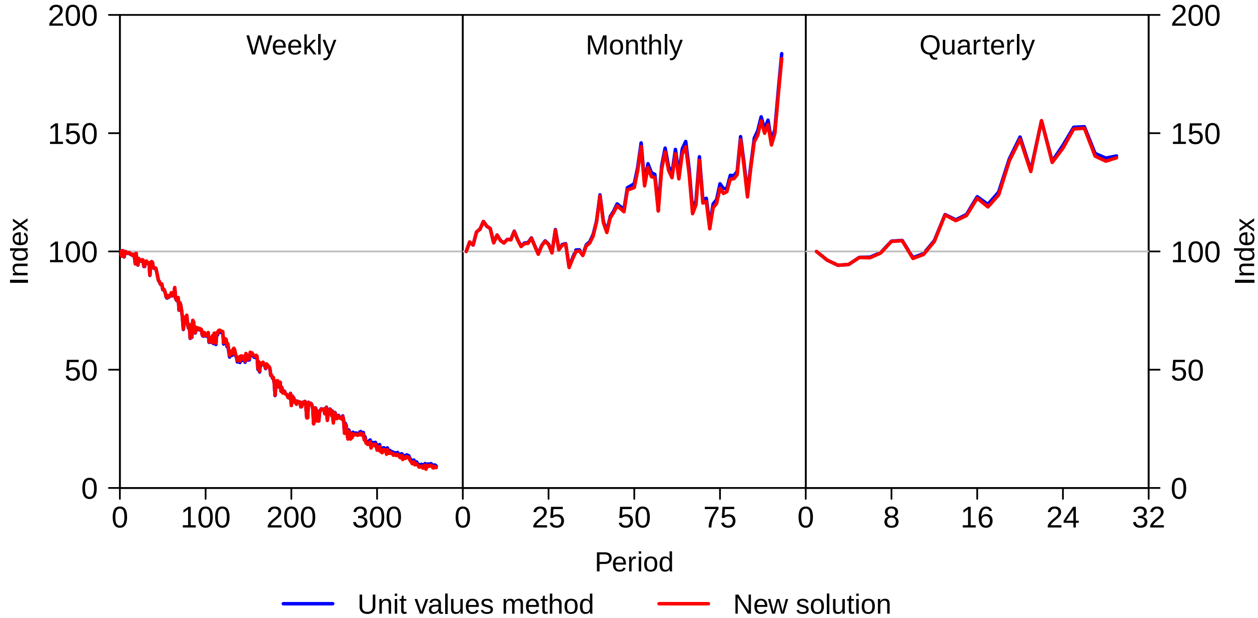

Figure 2: Törnqvist price indexes for laundry detergent

Notes: In scope are detergents sold at Dominick's, a now-defunct Chicago-area grocery store. The unit values method is standard. The new solution is described in the body text. Despite the extreme chain drift in the weekly indexes, the two methods give almost identical results.
Sources: Author's calculations; James M Kilts Center (University of Chicago Booth School of Business)

A.7 Exotic functional forms for hedonic indexes (ONLINE ONLY)

In this section I estimate a simple hedonic price index based on an arithmetic mean. The estimates are very close to those from the equivalent geometric mean index, but they are also more difficult to obtain. So departing from the geometric mean will rarely be worthwhile.

The estimated index is a chained monthly hedonic price index for detached houses in Sydney, covering the period from January 2007 to December 2011. The index thus comes from 47 regressions, one for each chain link. For each of those links I use a simplified definition of quality, $quality_{it} \equiv beds^{\beta_1} area^{\beta_2}$, where *beds* is number of bedrooms, *area* is land size in square metres, and the $\beta_{k\in\{1,2\}}$ are parameters to be estimated. I set $f(\cdot) = (\cdot)$ and, again for simplicity, $w_{it} = 1$.

To estimate the index I adapt the method of moments technique from Hajargasht & Rao (2019).[16] Focus is on the estimated fixed effects for t, which form the basis of each chain link. To avoid complexity that is unnecessary for my purposes I do not seasonally adjust. The data come from Australian Property Monitors (APM).

Figure 3 shows two versions of the index, differing in the algorithm used to solve the moment conditions (both in Chaussé 2010). The indexes are close to one another, and the one using the Fletcher (1970) algorithm is indistinguishable from the equivalent geometric mean index (not plotted). Both algorithms require starting guesses for parameter values, for which I use the analytical estimates from the geometric mean index. Using other starting values from the same broad vicinity produces identical estimates in this case. But using naïve starting values, like a vector of ones, produces nonsense. Such complexities will make these indexes less attractive for statistical agencies.

The script used to produce these figures is in the supporting material online, although the data are not, on account of being proprietary.

A.7.1 Copyright and Disclaimer Notice

The Sydney property price data used in this appendix is sourced from Australian Property Monitors Pty Limited ACN 061 438 006 of 10 Harris Street Pyrmont NSW 2009 (P: 1 300 655 177). In providing these data, Australian Property Monitors relies upon information supplied by a number of external sources (including the governmental authorities referred to below). These data are supplied on the basis that while Australian Property Monitors believes all the information provided will be correct at the time of publication, it does not

[16] I copy the procedure for what Table 3 of my Appendix describes as their "type 2" index. I just swap the product fixed effects with my hedonic definition of quality. A straight non-linear least squares procedure is inappropriate because the estimates of the quality parameters are more heavily influenced by periods with higher price levels (and hence higher measured volatility). This imbalance is problematic when the chosen functional form is a blatant approximation.

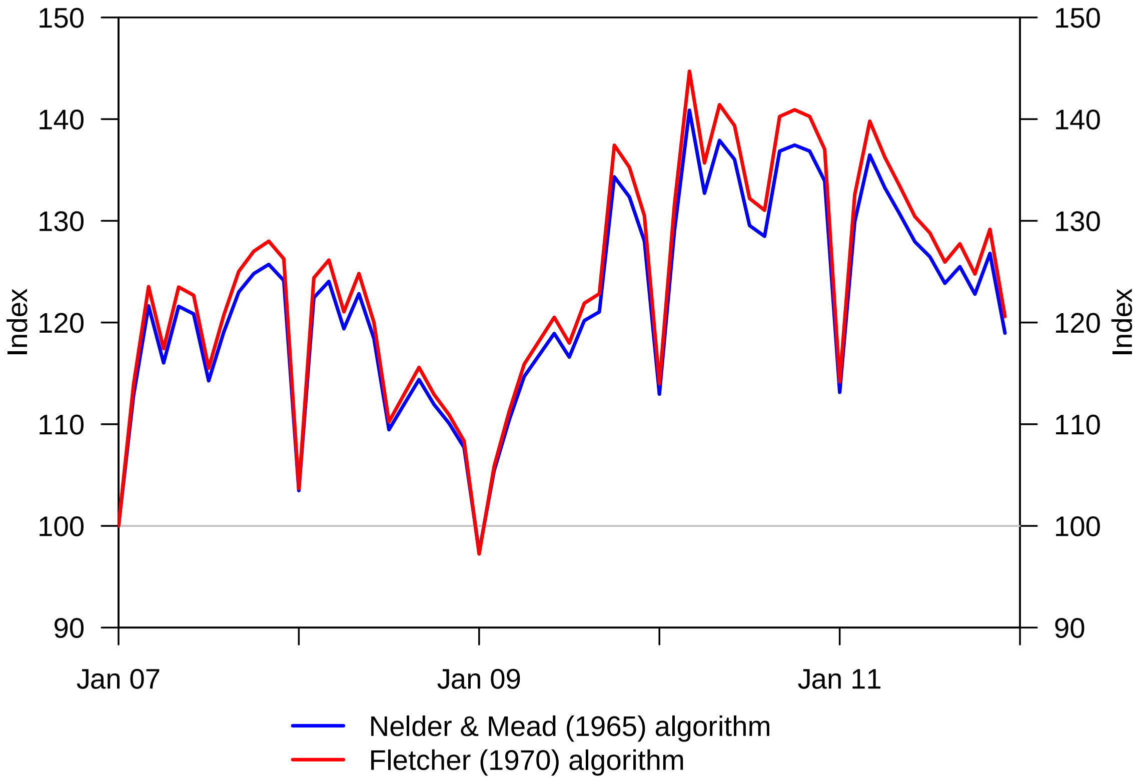

Figure 3: Hedonic price indexes (arithmetic) for detached houses in Sydney

Notes: The series are not seasonally adjusted. The algorithm from Fletcher (1970) produces an index that is indistinguishable from the equivalent geometric hedonic index (not plotted). The differences between popular algorithms, though small, is an unattractive feature of using arithmetic means for hedonic indexes.
Sources: Author's calculations; APM

warrant its accuracy or completeness and to the full extent allowed by law excludes liability in contract, tort or otherwise, for any loss or damage sustained by you, or by any other person or body corporate arising from or in connection with the supply or use of the whole or any part of the information in this publication through any cause whatsoever and limits any liability it may have to the amount paid to the Publisher for the supply of such information.

The data contain property sales information provided under licence from the Department of Finance and Services, Land and Property Information.

References

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J. & Li, H. (2019), 'Missing growth from creative destruction', *The American Economic Review* **109**(8), 2795–2822.

Balk, B. M. (2008), *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*, Cambridge University Press, Cambridge.

Banerjee, K. (1983), 'On the existence of infinitely many ideal log-change index numbers associated with the CES preference ordering', *Statistische Hefte* **24**(1), 141–148.

Berndt, E. R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading.

Berndt, E. R. & Rappaport, N. J. (2001), 'Price and quality of desktop and mobile personal computers: a quarter-century historical overview', *The American Economic Review* **91**(2), 268–273.

Berry, S., Levinsohn, J. & Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**(4), 841–890.

Brachinger, H. W., Beer, M. & Schöni, O. (2018), 'A formal framework for hedonic elementary price indices', *AStA Advances in Statistical Analysis* **102**(1), 67–93.

Bryan, M. F. & Cecchetti, S. G. (1994), Measuring core inflation, *in* N. Mankiw, ed., 'Monetary policy', Vol. 29 of *NBER Stuides in Business Cycles*, The University of Chicago Press, Chicago, pp. 195–215.

Chaussé, P. (2010), 'Computing generalized method of moments and generalized empirical likelihood with R', *Journal of Statistical Software* **34**(11), 1–35.
URL: *http://www.jstatsoft.org/v34/i11/*

Clements, K. W. & Izan, H. Y. (1981), 'A note on estimating Divisia index numbers', *International Economic Review* **22**(3), 745–747.

Clements, K. W. & Izan, H. Y. (1987), 'The measurement of inflation: a stochastic approach', *Journal of Business & Economic Statistics* **5**(3), 339–350.

Court, A. T. (1939), Hedonic price indexes with automotive examples, *in* 'The Dynamics of Automobile Demand', General Motors Corporation, New York, pp. 98–117.

Davies, G. R. (1924), 'The problem of a standard index number formula', *Journal of the American Statistical Association* **19**(146), 180–188.

Davies, G. R. (1932), 'Index numbers in mathematical economics', *Journal of the American Statistical Association* **27**(177A), 58–64.

de Carvalho, M. (2016), 'Mean, what do you mean?', *The American Statistician* **70**(3), 270–274.

de Haan, J. (2004), Hedonic regression: the time dummy index as a special case of the imputation Törnqvist index, *in* '8th Ottawa Group Meeting, Helsinki, 23-25 August'.

de Haan, J. & Krsinich, F. (2018), 'Time dummy hedonic and quality-adjusted unit value indexes: do they really differ?', *Review of Income and Wealth* **64**(4), 757–776.

de Haan, J. & Van der Grient, H. A. (2011), 'Eliminating chain drift in price indexes based on scanner data', *Journal of Econometrics* **161**(1), 36–46.

Diewert, W. E. (1976), 'Exact and superlative index numbers', *Journal of Econometrics* **4**(2), 115–145.

Diewert, W. E. (1981), The economic theory of index numbers: a survey, *in* A. Deaton, ed., 'Essays in the Theory of Consumer Behavior: In Honour of Sir Richard Stone', Cambridge University Press, Cambridge, pp. 163–208.

Diewert, W. E. (1992), 'Fisher ideal output, input, and productivity indexes revisited', *Journal of Productivity Analysis* **3**(3), 211–248.

Diewert, W. E. (2005a), 'Adjacent period dummy variable hedonic regressions and bilateral index number theory', *Annales d'Economie et de Statistique* **79/80**, 759–786.

Diewert, W. E. (2005b), 'Weighted country product dummy variable regressions and index number formulae', *Review of Income and Wealth* **51**, 561–570.

Diewert, W. E. (2010), On the stochastic approach to index numbers, *in* W. E. Diewert, B. M. Balk, D. Fixler, K. J. Fox & A. O. Nakamura, eds, 'Price and Productivity Measurement: Volume 6 - Index Number Theory', Tafford Press, Bloomington, pp. 235–62.

Diewert, W. E., Heravi, S. & Silver, M. (2009), Hedonic imputation versus time dummy hedonic indexes, *in* W. E. Diewert, J. S. Greenlees & C. R. Hulten, eds, 'Price Index Concepts and Measurement', Vol. 70 of *NBER Studies in Income and Wealth*, University of Chicago Press, Chicago, pp. 161–196.

Divisia, F. (1926), 'L'indice monétaire et la théorie de la monnaie (the monetary index and the theory of currency (conclusion))', *Revue d'Economie Politique* **40**(1), 49–81.

Feenstra, R. C. (1995), 'Exact hedonic price indexes', *Review of Economics and Statistics* **77**(4), 634–653.

Fisher, I. (1922), *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*, Houghton Mifflin Company, Boston.

Fletcher, R. (1970), 'A new approach to variable metric algorithms', *The Computer Journal* **13**(3), 317–322.

Fox, K. J. & Syed, I. A. (2016), 'Price discounts and the measurement of inflation', *Journal of Econometrics* **191**(2), 398–406.

Geary, R. C. (1958), 'A note on the comparison of exchange rates and purchasing power between countries', *Journal of the Royal Statistical Society. Series A (General)* **121**(1), 97–99.

Gorajek, A. (2019), The well-meaning economist, Discussion Paper 19-04, Reserve Bank of Australia. Accessed on 25 July 2019.
URL: *https://www.rba.gov.au/publications/rdp/2019/2019-04.html*

Griliches, Z. (1971), Hedonic price indexes revisited, *in* Z. Griliches, ed., 'Price indexes and quality change: studies in new methods of measurement', Massachusetts, Harvard University Press, pp. 3–15.

Hajargasht, G. & Rao, D. P. (2010), 'Stochastic approach to index numbers for multilateral price comparisons and their standard errors', *Review of Income and Wealth* **56**(S1), S32–S58.

Hajargasht, G. & Rao, D. P. (2019), 'Multilateral index number systems for international price comparisons: properties, existence and uniqueness', *Journal of Mathematical Economics* .

Hansen, B. E. (2021), 'Econometrics', Web publication of draft textbook manuscript. Accessed on 9 January 2021.
URL: *https://www.ssc.wisc.edu/ bhansen/econometrics/Econometrics.pdf*

Heravi, S. & Silver, M. (2007), Different approaches to estimating hedonic indexes, *in* 'Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches', University of Chicago Press, pp. 235–268.

Hill, R. J. (1997), 'A taxonomy of multilateral methods for making international comparisons of prices and quantities', *Review of Income and Wealth* **43**(1), 49–69.

Iklé, D. M. (1972), 'A new approach to the index number problem', *The Quarterly Journal of Economics* **86**(2), 188–211.

International Labour Office, International Monetary Fund, Organisaiton for Economic Co-operation and Development, European Union, United Nations & World Bank (2020), *Consumer Price Index Manual: Concepts and Methods*, International Monetary Fund, Washington, DC.

International Labour Organisation, International Monetary Fund, Organisaiton for Economic Co-operation and Development, United Nations Economic Comission for Europe & The World Bank (2004), *Producer Price Index Manual: Theory and Practice*, International Monetary Fund, Washington, DC.

Ivancic, L., Diewert, W. E. & Fox, K. J. (2011), 'Scanner data, time aggregation and the construction of price indexes', *Journal of Econometrics* **161**(1), 24–35.

Jevons, W. S. (1863), *A Serious Fall in the Value of Gold Ascertained and its Social Effects Set Forth*, Edward Stanford, London.

Keynes, J. M. (1930), *A Treatise on Money: Volume 1 The Pure Theory of Money*, MacMillan and Co. Limited, London.

Khamis, S. H. (1972), 'A new system of index numbers for national and international purposes', *Journal of the Royal Statistical Society. Series A (General)* **135**(1), 96–121.

Kloek, T. & De Wit, G. (1961), 'Best linear and best linear unbiased index numbers', *Econometrica* **29**(4), 602–616.

Koenker, R. & Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Konüs, A. A. (1939), 'The problem of the true index of the cost of living', *Econometrica* **7**(1), 10–29.

Kroese, D. P., Taimre, T. & Botev, Z. I. (2011), *Handbook of Monte Carlo Methods*, John Wiley & Sons.

Kurtzon, G. (2020), The problem with normalizing preferences that change in a cost-of-living index, Discussion Paper 534, US Bureau of Labor Statistics. Accessed on 10 February 2021.
URL: *https://www.bls.gov/osmr/research-papers/2020/ec200160.htm*

Lehr, J. (1885), *Beiträge zur Statistik der Preise: Insbesondere des Geldes und des Holzes [Contributions to Price Statistics: In Particular on Money and Timber]*, J.D. Sauerländer, Frankfurt am Main.

Lloyd, P. J. (1975), 'Substitution effects and biases in nontrue price indices', *The American Economic Review* **65**(3), 301–313.

Machado, J. A. & Santos Silva, J. (2006), 'A note on identification with averaged data', *Econometric Theory* **22**(3), 537–541.

Manski, C. F. (2015), 'Communicating uncertainty in official economic statistics: an appraisal fifty years after Morgenstern', *Journal of Economic Literature* **53**(3), 631–53.

Montgomery, J. K. (1937), *The Mathematical Problem of the Price Index*, P.S. King & Son, London.

Moulton, B. (2018), The measurement of output, prices, and productivity: what's changed since the Boskin commission, Report, Hutchins Center on Fiscal and Monetary Policy, Brookings Institution. Accessed on 14 July 2019.
URL: *https://www.brookings.edu/research/the-measurement-of-output-prices-and-productivity*

Muliere, P. & Parmigiani, G. (1993), 'Utility and means in the 1930s', *Statistical Science* **8**(4), 421–432.

National Research Council (2002), *At What Price? Conceptualizing and Measuring Cost-of-living and Price Indexes*, National Academies Press, Washington DC.

Pakes, A. (2003), 'A reconsideration of hedonic price indexes with an application to PC's', *The American Economic Review* **93**(5), 1578–1596.

Persons, W. M. (1928), 'The effect of correlation between weights and relatives in the construction of index numbers', *The Review of Economic Statistics* **10**(2), 80–107.

Rao, D. P. (1990), A system of log-change index numbers for multilateral comparisons, *in* J. Salazar-Carrillo & D. P. Rao, eds, 'Comparisons of Prices and Real Products in Latin America', Vol. 194 of *Contributions to Economic Analysis*, Elsevier Science Publishers B.V., Amsterdam, pp. 127–139.

Rao, D. P. & Hajargasht, G. (2016), 'Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP)', *Journal of Econometrics* **191**(2), 414–425.

Rao, D. P. & Selvanathan, E. (1992), 'Computation of standard errors for Geary-Khamis parities and international prices: a stochastic approach', *Journal of Business and Economic Statistics* **10**(1), 109–115.

Redding, S. J. & Weinstein, D. E. (2020), 'Measuring aggregate price indices with taste shocks: theory and evidence for CES preferences', *The Quarterly Journal of Economics* **135**(1), 503–560.

Reis, H. J. & Santos Silva, J. M. C. (2006), 'Hedonic price indexes for new passenger cars in Portugal (1997–2001)', *Economic Modelling* **23**(6), 890–908.

Rosen, S. (1974), 'Hedonic prices and implicit markets: product differentiation in pure competition', *Journal of Political Economy* **82**(1), 34–55.

Sato, K. (1974), 'Ideal index numbers that almost satisfy the factor reversal test', *The Review of Economics and Statistics* **56**(4), 549–552.

Sato, K. (1976), 'The ideal log-change index number', *The Review of Economics and Statistics* **58**(2), 223–228.

Selvanathan, E. & Rao, D. P. (1994), *Index Numbers: A Stochastic Approach*, The Macmillan Press Ltd, London.

Stoevska, V. (2008), Consumer price index manual: theory and practice, report on the 2007 world-wide survey on the usefulness of the CPI manual, Information paper, International Labour Organisation. Accessed on 13 April 2018.
URL: *http://www.ilo.org/public/english/bureau/stat/download/cpi/surv-ey.pdf*

Stuvel, G. (1957), 'A new index number formula', *Econometrica* **25**(1), 123–131.

Theil, H. (1960), 'Best linear index numbers of prices and quantities', *Econometrica* **28**(2), 464–480.

Törnqvist, L. (1936), 'The Bank of Finland's consumption price index', *Bank of Finland Monthly Bulletin* **16**(10), 1–8.

Triplett, J. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD Publishing, Paris.

Vartia, Y. O. (1976), 'Ideal log-change index numbers', *Scandinavian Journal of Statistics* **3**(3), 121–126.

Voltaire, K. & Stack, E. J. (1980), 'A Divisia version of the country-product-dummy method', *Economics Letters* **5**(1), 97–99.

von Auer, L. (2014), 'The generalized unit value index family', *Review of Income and Wealth* **60**(4), 843–861.

Walsh, C. M. (1901), *The Measurement of General Exchange-Value*, The Macmillan Company, New York.

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, MIT press, Cambridge.

World Bank (2013), *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program*, World Bank, Washington, DC.