Accounting for variety

Winkler, Julian

University of Oxford

24 May 2022

Accounting for variety*

Julian Winkler†[1,2]

[1]*Department of Economics, University of Oxford*
[2]*Institute of New Economic Thinking at the Oxford Martin School*

May 24, 2022

Abstract

I introduce a general method to account for the distribution of underlying components (variety) of an aggregate quantity, using the notion of entropy. This accounting decomposition enables a number of insightful applications for index numbers in economics. The cross-entropy of GDP with respect to a benchmark captures the change in its distribution, and thus how well this benchmark matches data for price and volume indices across time. This 'error' changes demonstrably over time. Accounting of variety also lends itself to an decomposition of labour productivity growth by a technology component (how many more 'average' goods are produced per unit of labor?), an allocation component (does the distribution of labor inputs converge to the distribution of outputs?), and cost disease (does the distribution of expenditures diverge from the distribution of outputs?).

*Acknowledgements: I thank Rick Van der Ploeg, François Lafond and Ian Crawford for valuable comments. I am grateful to Baillie Gifford and the Institute for New Economic Thinking at the Oxford Martin School for funding this research, and Valentina Semenova for sharing her data.

†julian.winkler@economics.ox.ac.uk

1 Introduction

Real output of goods and services is the accepted measure of historical living standards. To track its evolution, statisticians count growth in expenditures, also 'nominal' output, hoping to separate growth in price (inflation) versus real output:

$$\underbrace{\text{Real output growth}}_{\text{extrapolated}} = \underbrace{\text{Nominal output growth}}_{\text{accounting data}} - \underbrace{\text{Price Inflation}}_{\text{survey data}}.$$

Nominal output is neatly observed in currency units, but *what* exactly should we measure the price, and real output, of? Choosing a measure for price requires aggregating prices from dozens of industries, hundreds of product types, and millions of firms. In abstract terms, 'what' we pay for at any given point in time is in the distribution of nominal output among a number of 'types' of output: industries, products, firms, etc.

This paper notes that even if the question of 'what' we want to measure is settled, 'what' we pay for changes over time. For example, we know the distribution of US nominal output in 2000 to rely more on service and IT industries compared to 1950. If we are interested in measuring historical real output growth from the perspective of the average firm in 2000, more weight should be placed on inflation for service and IT industries in the 1950's, even if they constituted a smaller share of nominal output. The result is a worse approximation of real output growth for the average firm in 1950.

It follows that the importance placed on certain goods should change with the distribution of nominal output as preferences evolve. However, in a recent study, Baqaee & Burstein (2021) insist on fixing weights, to the extent that they change due to income effects or taste shocks, when measuring economic welfare. The important contribution is their challenge to what we should interpret as a historically consistent, and intuitive, measure for living standards. The composition of expenditures they seek to deflate may not agree with the weights assigned to price components, which this paper shows to hold significant practical implications.

One illustrative thought experiment demonstrating the importance of a changing output distribution at any level of aggregation is Hulten's Paradox. The ICT revolution in the 1990's sparked a discussion on measuring welfare improvements from the changing quality of products, in addition to their real output. Besides practical difficulties in implementing such 'hedonic' price indices, Hulten (1997), in response to Nordhaus's (1997) claim that the hedonic price of light dropped faster than its traditional counterpart, famously cautions the use of quality adjustments since they likely overstate actual welfare gains. A passage from Hulten's (1997) comment, evaluating Nordhaus's (1997) revised price deflator, provides the key insight:

> [...]a person possessing the average disposable income in America today should be willing to accept a massive reduction in spending power – from \$17,200 to the \$90-430 range – in order to avoid being sent back in time to an equivalent status in colonial America. Alternatively, it suggests that the average colonial should

> prefer living in the America of today, with as little as $90 per year, to staying put in the late eighteenth century.

The contentious point in Hulten's quote is that a basket of goods for the colonial American evolved very differently from a representative basket of goods consumed today. The $90 today price a different 'representative' unit than $90 in colonial America. An individual today would have indeed turned out quite poor in colonial America, if the representative basket of goods was substantially more scarce compared to the basket of a colonial American. Sending him back in time requires centuries of devolving preferences and substitutions. This substitution also entails a scarcer basket of goods for a colonial American, today.

Section 2 formalises this issue using index number theory, and introduces an extra entropy term, alongside price and real output, which encapsulated the shift in consumption patterns. Entropy is used in many fields as a measure of 'observational variety', and therefore lends an elegant application in the present economic context. The result emerges naturally from Jensen's Inequality, but also benefits from deep intuition; indices aggregate across a variety, so a change in variety hurts the 'representativeness' of an index.

The entropy term captures a bias in real output when extrapolated from the difference in nominal output and price inflation. Furthermore, this bias travels in one direction with time, providing an intriguing prediction: the representativeness of a basket of goods from the past should worsen, whereas it improves for today's basket. In practice, this means any pair of price and real output indices that favour recent consumption bundles will overestimate the price and real output growth rates of older bundles. This finding can be used to evaluate the bias in popular indexation methods. For example, Divisia indices update product weights in each period, which achieves the best result in terms of minimising the influence of the cross-entropy bias.

This result applies to the any aggregate index. One application is in measuring the reallocation effect studied in the growth accounting field, whereby resources move between high and low productivity industries. In this context, the entropy of employment by industry contributes to a labour productivity decomposition, to the degree that the allocation of employment diverges from sources of production.

Section 3 measures the bias from cross-entropy in three applications. The tests the effect in a transparent manner, using industry price and output data from US KLEMS. Forming production profiles from 65 Standard Industry Classification (SIC) industries, the real output growth rate experienced by the 1947 bundle is over 1pp smaller than that experienced by the 2014 bundle. Without correcting for cross-entropy, this means that we would overstate today's living standards by around 70% compared to those of 50 years ago when using consumption patterns from today, as would be the case if data on industry-level real output was unavailable. Overall, the industries that made up larger shares of nominal output in 1947 experienced both lower inflation and real output growth rates. It follows that nominal output grew more uniform between industries, as reflected by an increasing rate of entropy. A popular view is that the economic structure of the US shifted out of agriculture and into professional or health services, due to some type of 'cost disease'.

Two further applications are instructive for the wider applicability of entropy to economic questions. First, I reproduce an official Consumer Price Index (CPI) for countries in which statistical agencies publish product-level price data for around 200 categories. One interesting finding is that the contribution to entropy at this granular level is similar to the contribution at a higher industry level. Second, I extend the output exercise by including a labour index, which yields a labour productivity decomposition between i) the productivity growth rate experienced by the average industry, ii) the divergence of expenditures from important industries, and iii) the convergence of the distribution, by industry, of labour inputs to the importance of any given industry assigned by the weighting scheme. Using EU KLEMS data, I find that a significant part of the post-2005 slowdown in labour productivity growth in Germany can be attributed to a slowdown in the allocation of labour inputs.

Related literature: A number of key arguments characterise the persistent debate on what real output should measure. Complications arise when new products enter, or old products leave, the market: Aghion et al. (2019) estimate that imputations for inflation of products subject to creative destruction leads to an overestimation of the true inflation rate by an average of 0.5%. Most of this stems from hotel, restaurant and retail trade industries. There exists an active literature that seeks to interpret changes in product variety in terms of welfare (Redding & Weinstein 2019, Baqaee & Burstein 2021). The accounting framework I present is complementary to those efforts, but extends beyond the scope of output and welfare measurements.

Revisiting Hulten's Paradox for a moment, Gordon (2009) forces a point on CPI measurement, demonstrating that US women's apparel products experienced higher rates of inflation under a hedonic price model. To be clear: the entropy problem is distinct, in the sense that real output growth rates may appear unreasonably high regardless of downward biases in aggregate deflators. The cross-entropy term behaves in opposing directions for aggregate prices weighted to the start of the period, from which consumption baskets diverge, versus the end, towards which consumption baskets converge.

Section 2 presents the decomposition, and explains the role of entropy in the present context. Section 3 applies this decomposition to US KLEMS data to demonstrate the scale of the cross-entropy problem, and infers what the bias might be from official CPI weights. Section 4 concludes.

2 Indexing variety

In this section, I develop the accounting framework for the context of aggregate nominal output, or Gross Domestic Product (GDP), but it applies to any index that seeks to aggregate across a variety of types: employment, exchange rates, monetary aggregates, etc. National statistical agencies measure GDP and its composition between industries. The relationship between aggregate and industry nominal outputs, themselves the products of industry prices times real outputs, adheres to Definition 1.

Definition 1 (GDP). *GDP, Y, is the sum of N industry nominal outputs Y_i, which themselves are equal to the product of the industry-specific price level P_i and real output level Q_i:*

$$Y = \sum_{i=1}^{N} Y_i = \sum_{i=1}^{N} P_i Q_i. \tag{1}$$

Note that Definition 1 implies the existence of industry-level deflators P_i, which simplifies the exposition. When dealing with long-term economic data, statisticians typically observe prices from consumer surveys, then extrapolate an aggregate real output index from aggregate GDP. I make two assumptions throughout this paper. First, the number of types(industries) N is large. Second, GDP and industry prices P_i are always known.[1] Industry-specific nominal or real outputs are not necessarily observed.[2]

To track historical living standards, statisticians look to retrieve an index that aggregates real outputs across industries over time. Since these are unobserved, the next best thing is to construct an aggregate price index with which to deflate GDP. Definition 2 states the type of index studied throughout this paper.

Definition 2 (Price and real output indices). *The price(real output) index equals a geometric weighted average of industry prices(real outputs),*

$$\widetilde{P}_t = \Delta \sum_{i=1}^{N} \omega_{i,t} \log P_{i,t}, \quad \widetilde{Q}_t = \Delta \sum_{i=1}^{N} \omega_{i,t} \log Q_{i,t}, \tag{2}$$

where t denotes time, ω_i denotes a positive weight assigned to industry i such that $\sum_i \omega_i = 1$, and Δ is the difference operator, $\Delta x \equiv x_t - x_{t-1}$.

Industry weights ω_i belong to a $1 \times N$ vector $\mathbf{\Omega}$, which I refer to as an *indexation scheme*. This is distinct from the distribution of industry nominal output shares Y_i/Y, populating a $1 \times N$ vector $\mathbf{Y}$, which I refer to as *consumption patterns*.

An important remark is that Definition 2 restricts the scope of the paper to indices constructed from geometric averages. Examples are the Törnqvist, Sato-Vartia and Divisia indices, but alternatives exist, notably the Fisher, Laspeyres and Paasche indices (Törnqvist 1936, Sato 1976, Vartia 1976, Fisher 1922). A sprawling literature, duly explained by Eichhorn (1978), proposes sets of tests to evaluate which index is 'best'. The main contender to the type of index introduced by Definition 2 is Fisher's ideal index.[3]

[1] One complication, especially for longer time series, is that surveyed prices are often consumer prices, not producer prices. One symptom of this problem is that a price index used to deflate *domestic* output may include import prices. While this challenge is beyond the scope of the present paper, the assumption offers a best-case scenario in measuring historical living standards.

[2] Historically, the breakdown of GDP by industry is not always available, for two main reasons; i) tracking expenditures for granular product classifications, which can include hundreds of product types, is a significant practical challenge, and ii) data attrition increases the longer the time series, even for coarser industry aggregates.

[3] I briefly summarise Eichhorn (1978) on the trade-off between Fisher-type indices and Törnqvist-type indices, and its relevance to the present context. Fisher-type indices fail a *Base Test*, in that the total inflation rate between two periods is not equal to the sum of inflation rates in enclosed sub-periods, whereas Törnqvist-type indices fail a *Product Test*, in that the aggregate price and real output indices do not sum to the GDP index.

Indices of the type in Definition 2 produce a residual, because

$$\widetilde{P}_t + \widetilde{Q}_t = \Delta \log Y_t \quad \Longleftrightarrow \quad \Delta \log P_{i,t} = \Delta \log P_{j,t} \quad \text{and} \quad \Delta \log Q_{i,t} = \Delta \log Q_{j,t}, \quad \forall \quad i,j. \tag{3}$$

Unless price and quantity growth is exactly equal for each category i, the resulting indices will not sum to GDP growth. The key contribution of the present paper is to give this residual a name. This is desirable, because it yields an exact decomposition of GDP in an economically meaningful way, without resorting to assumptions required by complicated demand systems. Proposition 1 introduces this decomposition, applied to GDP growth.

Proposition 1 (Accounting for variety). *GDP can be decomposed exactly between a price index, real output index, and the cross-entropy of consumption patterns relative to the indexation scheme:*

$$\Delta \log Y_t = \underbrace{\Delta \sum_{i=1}^{N} \omega_{i,t} \log P_{i,t}}_{\text{Price inflation: } \widetilde{P}_t} + \overbrace{\underbrace{\Delta \sum_{i=1}^{N} \omega_{i,t} \log Q_{i,t}}_{\text{Real output growth: } \widetilde{Q}_t} + \underbrace{\Delta \sum_{i=1}^{N} \omega_{i,t} \log\left(\frac{Y_t}{Y_{i,t}}\right)}_{\text{Entropy change: } \Delta H(\mathbf{\Omega}_t, \mathbf{Y}_t)}}^{\text{Extrapolated output growth: } \widetilde{X}_t} . \tag{4}$$

Proof. The difference between GDP and an index of industry nominal outputs, following indexation scheme $\mathbf{\Omega}_t$, is

$$\log Y_t - \sum_{i=1}^{N} \omega_{i,t} \log Y_{i,t} = \sum_{i=1}^{N} \omega_{i,t} \log\left(\frac{Y_t}{Y_{i,t}}\right),$$

which is the cross-entropy of the indexation scheme $\mathbf{\Omega}_t$ and the consumption pattern $\mathbf{Y}_t$. Following convention, this term is written as $H(\mathbf{\Omega}_t, \mathbf{Y}_t)$. Substituting $\log Y_{i,t} = \log P_{i,t} + \log Q_{i,t}$, which follows from Definition 1, and re-arranging yields the desired result.

□

There are two immediate observations on the residual, termed 'cross-entropy', in Proposition 1. First, cross-entropy is identical to the negative log-likelihood of the nominal output data, $\mathbf{Y}_t$, relative to $\mathbf{\Omega}_t$. Therefore, minimising cross-entropy is similar to maximising the log-likelihood of all $Y_{i,t}$, given that we assume weights $\omega_{i,t}$. The second interpretation is that of entropy itself: cross-entropy quantifies the information required to encode Y_t, given that the encoding scheme is optimised for $\mathbf{\Omega}_t$. It is, loosely, a measure of the variation in $\mathbf{Y}_t$ explained by the variation in $\mathbf{\Omega}_t$.

In practice, Proposition 1 suggests that statisticians have some freedom in choosing an indexation scheme that is economically meaningful, in that they produce a model $\mathbf{\Omega}_t$ to explain variation in expenditures across industries $\mathbf{Y}_t$. This comes at the price of some information loss, to the extent that the model does not reproduce true consumption patterns.

The failure in indices of the type in Definition 2 for the product test in Eq. 3 follows from Proposition 1. Corollary 1.1 finds that no such index can perfectly satisfy the product test: there always remains a positive residual from cross-entropy, which is bounded by the entropy of aggregate GDP.

Corollary 1.1. *The choice of indexation scheme* $\mathbf{\Omega}_t$ *which minimises* $H(\mathbf{\Omega}_t, \mathbf{Y}_t)$ *is*

$$\omega_{i,t} = \frac{Y_{i,t}}{Y_t} \quad \forall i. \tag{5}$$

Under this indexation scheme, cross-entropy can be written as

$$H(\mathbf{\Omega}_t, \mathbf{Y}_t) = \sum_{i=1}^{N} \frac{Y_{i,t}}{Y_t} \log\left(\frac{Y_t}{Y_{i,t}}\right) = H(\mathbf{Y}_t), \tag{6}$$

which is Shannon's (1948) definition of entropy, applied to aggregate nominal output Y_t*.*

Proof. Follows from the definition of cross-entropy (Cover & Thomas 2005). □

From Corollary 1.1, the cross-entropy residual can be minimised by setting the indexation scheme equal to the distribution of GDP. The existence of any residual is immediate from Eq. 5, which is a standard representation of Jensen's Inequality. When the indexation scheme is chosen to be equal to the distribution of GDP among industries, the cross-entropy term reduces to the entropy of GDP, $H(\mathbf{Y}_t)$, as defined by Shannon (1948). This index is typically referred to as the Divisia index.

The surprising implication of Corollary 1.1 is that the fidelity of any pair of price and real output indices to the product test has a hard boundary equal to the entropy of GDP. This has particular consequences for long time series, where the entropy of GDP itself might change over time. This point is developed further below.

2.1 Intermezzo: what is entropy?

Here, I illustrate what the entropy of GDP measures. Formally, recall that the market economy is accounted for by the transactions across N goods in a given period of time, for which the sum is GDP. The number of possible arrangements for those transactions is given by the multinomial coefficient:

$$\binom{Y!}{Y_1!, Y_2!, \ldots, Y_N!} = \frac{Y!}{Y_1! Y_2! \ldots Y_N!} = \frac{Y!}{(Yy_1)!(Yy_2)! \ldots (Yy_N)!},$$

where $y_i = Y_i/Y$. This is simplified, under large N, using the Stirling approximation:

$$\log\left[\frac{Y!}{(Yy_1)!(Yy_2)! \ldots (Yy_N)!}\right] \approx Y \sum_{i=1}^{N} y_i \log\left(\frac{1}{y_i}\right), \tag{7}$$

which corresponds to Shannon's (1948) entropy. To better interpret this term, assume that, on a given day, €10 of fruit are traded on a market consisting of two types, apples and coconuts. If spending is uniform, the number of possible allocations is

$$\frac{10!}{5!5!} = 252.$$

On the other extreme, if all but one Euro is spent on one fruit alone, the number of possible arrangements is

$$\frac{10!}{9!1!} = 10.$$

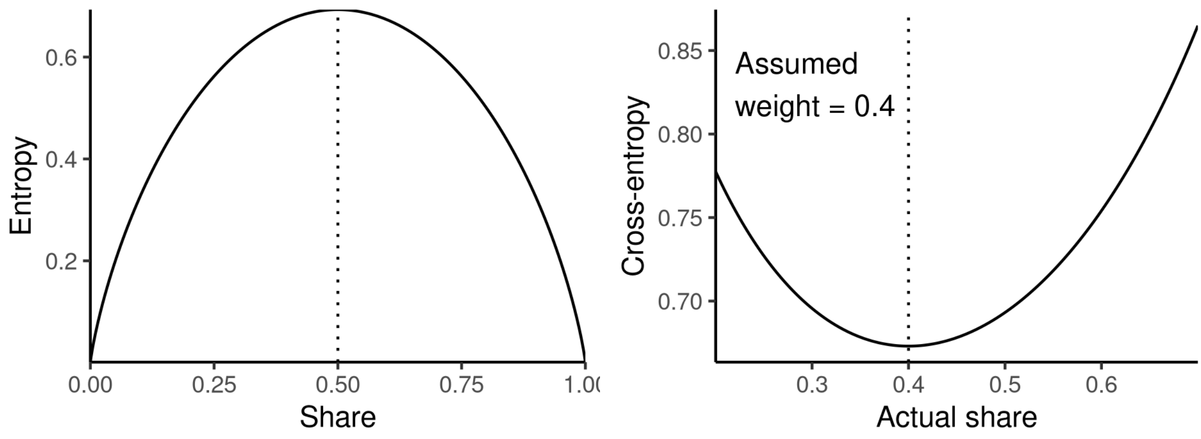

Figure 1: **Plotting entropy for two industries**: total entropy (Left) is maximal at the dotted line, where the share of each industry is equal at 1/2. By fixing a weight (Right) at the dotted line, a change in the actual share increases cross-entropy.

This is why entropy is often described as a measure of uncertainty, or *observational variety*: from the consumer's perspective, there are many more ways to spend an endowment equally among options, rather than prioritising some ahead of others.[4] An additional dollar has many possible ways of being spent in a uniform economy, but fewer in a concentrated economy.

Returning to the problem at hand, a weighting scheme $\mathbf{\Omega} \overset{d}{=} \mathbf{Y}$ yields an indexation error equal to the entropy of GDP. Therefore, the indexation error *worsens* as variety *increases*. In the example, a real output index will be more inaccurate when expenditures are similar between apples and coconuts, rather than concentrated in apples. This is visualised on the left of Figure 1: if all expenditures are on apples, then the growth rate of apples produced is a perfect measure for aggregate growth, and entropy is zero. Alternatively, if expenditures are allocated uniformly, the average growth rate is a worse approximation of the growth rates in either apples or coconuts. When national statisticians extrapolate an aggregate real output index from a time series of industry GDP and prices, it will invariably be progressively *more* biased *upwards* if variety in GDP increases. This is is particularly interesting for fixed weights, visualised on the right of Figure 1; the cross-entropy between the assumed weight versus the actual share can change over time, even if entropy itself is already low.

I write 'economic' variety because this result naturally extends beyond industry-level aggregation, right down to the product level. It is not hard to imagine that variety today is orders of magnitudes larger than a century ago. In fact, much of this is reflected in the constant revisions and expansions of industry classifiers: what products are similar enough that we can assume a common price level? The answer inevitably evolves with innovation and creative destruction, in the spirit of Aghion & Howitt (1992), Aghion et al. (2019).

[4] Two technical remarks: i) observational zeros do not appear ($0 \log 0 = 0$), and ii) each Euro is treated equally.

2.2 Historical inflation, real output and entropy

Returning to the problem at hand, the goal of index number theory is to infer how real output evolves over time, as a proxy for historical living standards. The logic behind Eq. 4 is as follows. Given that national accountants observe prices and aggregate GDP, the standard is to extrapolate the aggregate real output index, $\widetilde{X}_t$, from the difference in GDP growth rate, $\Delta \log Y_t$, and the choice price index, $\widetilde{P}_t$. However, this value is not equal to the true real output index, $\widetilde{Q}_t$, which the national accountant would find if industry-level output was observed. Instead, it includes cross-entropy, $\Delta H(\mathbf{\Omega}_t, \mathbf{Y}_t)$, and is thus a biased version of the actual real output growth index, to the extent that growth in cross-entropy differs from zero.

If weights are available and updated continuously, the full residual from cross-entropy equals the change in entropy of GDP. The residual is zero if the growth rate of nominal output is equal between all industries. The good news is that any residual is therefore small as long as the distribution of nominal output remains relatively stable. There are several instances where this might not be the case. For example, developing and middle income countries experience unbalanced industry growth. Other types of indices might suffer as well; is employment growth today well represented by employment growth experienced by blacksmiths and coopers? A century ago, it was.

In the present context, this section reaches one important conclusion. Real output indices of the type defined in Definition 2 are biased if extrapolated from GDP using a counterpart price index, to the degree that the indexation scheme differs from consumption patterns. In addition, this bias from cross-entropy in the indexation scheme and consumption patterns enters real output growth indices by at least the amount that the entropy of GDP changes. This bias is larger if the indexation scheme does not match consumption patterns, as is the case for common, economically-motivated real output indices.

3 Indices, revisited

3.1 Real output: Hulten's Paradox in KLEMS data

What does Proposition 1 imply for the measurement of living standards in the long run? To demonstrate the role of changing consumption patterns, I revisit Hulten's Paradox. Hulten's (1997) established the paradox response to Nordhaus's (1997) claim that prices grew slower than commonly accepted. He noted that the revision implies faster volume growth: when extrapolated back in time, this correction in prices leaves our ancestors with very little income in real terms. A paradox manifests in the observation that these ancestors were demonstrably capable to procure sustainable lifestyles.

The exchange centred on pricing attributes for products (dis)appearing for the first time, but this paper takes issue with one particular step in Hulten's logic.[5] He asserts that if true

[5]Quality improvements, the like Nordhaus (1997) investigates, are indeed important in biasing inflation indices upwards in matched-model methods; the availability of characteristics is increasing with new prod-

price growth is overestimated, and volume growth underestimated, the average colonial and modern Americans are indifferent between the average living standard of colonial America, and that afforded with \$90-\$430 in 1997. However, not only is the budget of 90-430 1997 dollars fixed, but so is the 1997 assumed consumption basket. The decisions from both modern and colonial Americans to travel across time should not be equivalent, because they adhere to different consumption patterns. Goods consumed in 1997 were likely scarce in colonial America – think of the entertainment a new Game Boy offers – while goods consumed by colonial Americans are abundant in 1997. The modern CPI Hulten uses to deduce living standard improvements one hundred years ago is off the mark.

Table 1: **Yearly GDP growth with different industry compositions: US, 1947-2014**

	$\Delta \log Y_t$	$\widetilde{P}_t$	$\widetilde{Q}_t$	$\Delta H_t(\Omega, Q)$
Basket from 1947	6.33	3.17	2.40	0.76
Basket from 2014	6.33	3.53	3.48	-0.68
Difference	0.00	-0.36	-1.08	1.44
Törnqvist 1947&2014	6.33	3.35	2.94	0.04

Notes: This table decomposes average nominal GDP growth from 1947 to 2014 between the average growth in price, volumes, and change in cross-entropy. The weights, Ω, are denoted in the rows. All values denote log points. Data are made available by World Klems.

Is the bias in real output due to entropy significant in practice? Returning to Eq. 4, the cross-entropy term increases over time as real output growth diverges between industries from the perspective of a colonial American, even when keeping inflation rates equal across industries. From the perspective of the modern American, aggregate GDP appears to have grown substantially, but this is only due to a decline in cross-entropy: they happen to consume more goods that experienced higher real output growth rates.

In order to capture the order of magnitude of this difference, I reproduce an aggregate GDP deflator from US World KLEMS data.[6] These data cover 65 industries, from 1947 to 2014. The files provide yearly GDP by industry in current and constant prices. From this, I back out industry-specific price indices by dividing reported industry-level nominal gross output by its repective output in real quantities. Using these gross output deflators as P_i's, I can estimate each term in Eq. 4 using different weighting schemes. Specifically, in Table 1 I compare the composition of average, yearly GDP growth for a representative basket in 1947 – fixing weights to industry shares of GDP in 1947 – to a representative basket in 2014 – with weights fixed to industry shares in 2014.

The first result is that, as expected, real output growth rates are higher when indexing industry contributions to the distribution of nominal output from 2014, relative to 1947. This difference is significant, in the order of 1pp per year, on average. This translates to

ucts, but on paper their prices may appear the same. Gordon (2009) proposes that, just as matched models bias inflation upwards, so they may bias inflation downwards, with apparel prices as a case in point.

[6] http://www.worldklems.net/data.htm

overstated real welfare improvements of about 50% after seven decades, if consumption patterns in 2014 are used as the benchmark. The second, surprising outcome of this exercise is that aggregate deflators contribute *positively* to this gap: the average product in 1947 turned out *less* expensive than the average product in 2014. Therefore, a resolution to Hulten's Paradox may not lie at all in revising aggregate price indices upward by finding negative biases, but rather in accounting for the change in industrial variety of GDP.

How does this compare to measures of welfare motivated by economic theory? Since Diewert's (1976) seminal contribution, statistical agencies adopted the use of standard techniques in forming aggregate price indices, motivated by economic theory. One recommended method is the Törnqvist index, which adopts a similar form to Eq. 4. Specifically, it assigns weights equal to the average output shares between two periods. Diewert motivates the properties of this index, which he defines to be 'superlative', by demonstrating how it benchmarks the price to an assumed average utility between the two periods of choice. Practically, it is useful because it may cancel the cross-entropy term in Eq. 4, depending on how evenly the true weights change over time.

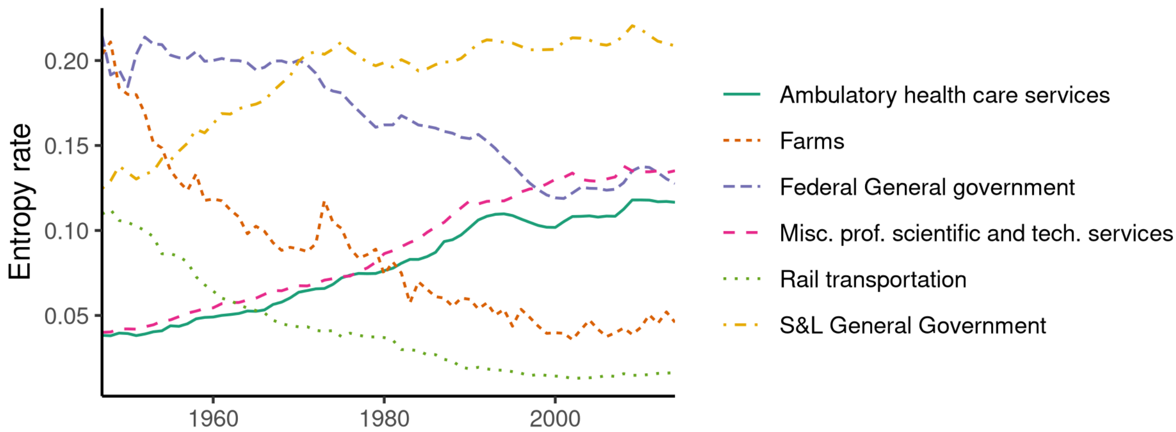

Figure 2: **Largest contributions to US entropy, 1947-2014**; this figure plots the three industries that have the three largest positive and negative contributions to the entropy of US GDP. Data are made available by World Klems.

The final line of Table 1 reports the contributions to GDP growth by adopting average weights between 1947 and 2014 as the benchmark. As expected, the loss of information by this index is substantially lower, and ever so slightly positive. This would suggest that industry shares evolved relatively smoothly during that period. Figure 2 helps understand where most of the change in US entropy comes from, by plotting the three largest positive and negative contributions to the overall entropy rate. The industries that emerge are well-known to those who study structural change of the US economy; the largest declining industries are agriculture, rail transport, but also output from federal government activities. On the other hand, the rise of health care and professional services, and S&L general government contributed most to an increasing entropy rate.

3.2 CPI: a role for granularity

Since the entropy of GDP clearly matters in measuring historical living standards, one wonders whether the 65 SIC industries offered by Worlds KLEMS paint an accurate picture. As mentioned in Section 1, data on disaggregated output, and thus weights, are hard to come by. Privacy protection rules also prevent publication of the most granular decomposition in industry output.

However, through remarkable efforts of statistical agencies in recent decades, data on components of consumption are more extensive, reliable and granular. On the product level, the BLS is tasked with assigning weights to products consumed by households for the purpose of estimating its CPI, which entails the use of surveys. The BLS publishes price indices for 182 products from 1998 to 2018, of which 10 have data for only a subset of years.[7] These prices are accompanied by official weights used to form aggregates, from surveys of what consumers buy for day-to-day living. After building a price index from these data, one can deflate any consumption series. I supplement the results from BLS data using the Harmonised Index of Consumer Prices (HICP) published by Eurostat, showing that the cross-entropy bias typically varies between 1-4pp per year across 33 European countries.

Section 2 established that this real output series is biased by the cross-entropy of CPI weights and the actual distribution of consumption. This bias is minimal when the weights are equal to the distribution of output, at which point it is equal to the entropy of consumption. Therefore, even though expenditure data is unavailable for the 182 products, the weights alone determine a 'best-case' outcome for this bias.

Table 2: **Yearly growth in consumer price and cross-entropy: US, 1998-2018**

	$\Delta \log \widetilde{P}_t$	$\Delta H(\mathbf{\Omega}_t)$
Basket from 1998	2.07	0.57
Basket from 2018	2.12	-0.48
Difference	-0.05	1.05
Törnqvist 1998&2018	2.08	0.06

Notes: BLS data on prices and weights for 182 products are used to form price indices in the first column, fixing weights to 1998 and 2018. The second column computes the average cross-entropy of those weights and the weights in all other years, denoted ω. All values denote log points. Data are made available by the BLS.

Table 2 computes this bias with CPI weights provided by the BLS. There are persistent differences between the CPI and the PPI, so these results do not compare directly to those of Table 1. However, the magnitude of the bias is surprisingly similar, at about 1pp per year. The bias for the Törnqvist index is also close, however small, at 0.06pp. At a rate of 1pp, deflating consumption using current expenditure patterns overstates improvements in living standards from fifty years ago by *at least* 65%. Another finding reproduced in Table 2 is the smaller inflation rate for the index based on 1998 consumption patterns.

[7] Omitting the incomplete series does not significantly alter results.

The European HICP, maintained by Eurostat, mimics the construction of the BLS' CPI, with the advantage of offering granular data for 33 European countries. Published data for the HICP include five levels of aggregation, with the 5-digit level counting up to 264 product categories. Decent coverage is available at that level for France, Lithuania and Slovenia. For the remaining countries, data is available at the 4-digit level, which includes up to 72 product categories. Generally, the price data span from 1996 to 2020, with some exceptions. Using those observations, I repeat the exercise from BLS CPI data for each European country. This will illustrate how varied the change in cross-entropy can be across different economies.

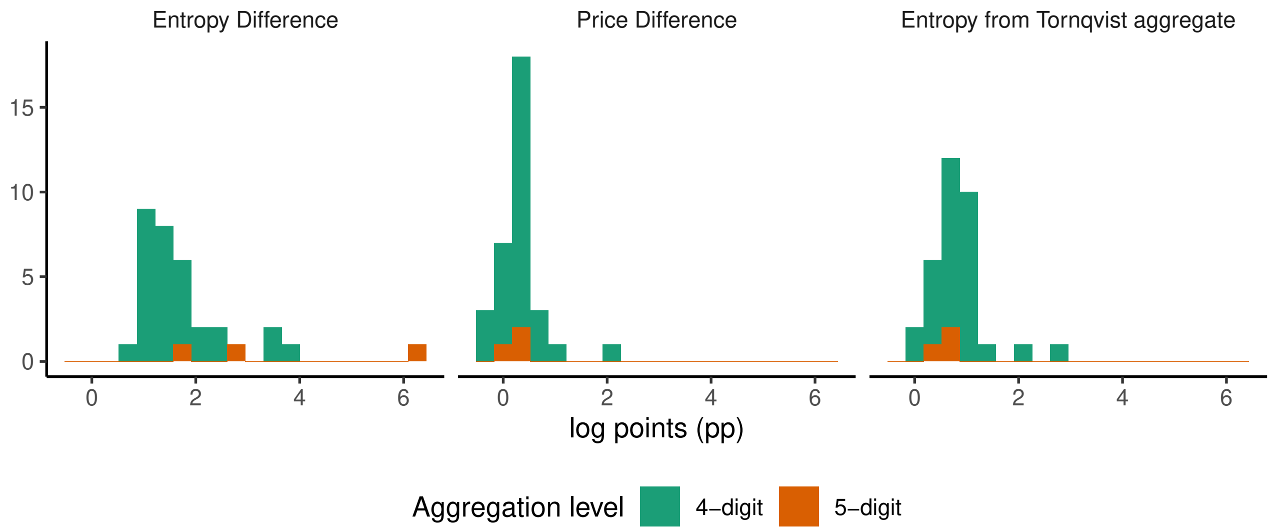

Figure 3: **Comparing price and entropy from first and last available weights in HICP data: 33 European countries, 1997-2020**;

By using weights from the first and final years available, Figure 3 reproduces the difference in average yearly changes in cross-entropy on the left, the difference in average yearly price growth rates in the middle, and the average yearly change in cross-entropy when using a Törnqvist average weight. These are comparable to row three, and the final column in row four, of Table 2.

The left chart in Figure 3 demonstrates that the average rate of information loss from cross-entropy for an aggregate real output index is between 1-2pp per year. These observations align closely to the 1.05pp found for US CPI data, but exhibit some variation, with 6.31pp an extreme case in Lithuania. It is not clear whether the level of aggregation changes has an obvious impact on cross-entropy estimates, although the number of countries publishing 5-digit level data is too low for a robust hypothesis test.

The second chart suggests that yearly price growth rates are higher when assigning weights from the first year of middle relative to the final year. This evidences some sort of substitution bias, by which consumption in later years favors products which grew relatively cheaper.

Finally, the right chart indicates that even Törnqvist aggregation, by which an average weight is derived from weights in the first and final years, can be subject to substantial bias from a change in cross-entropy. This suggests that the small result in US data, seen for industry deflators in Table 1 and in CPI weights in Table 2, may be a special case. In contrast, consumption patterns in other countries are subject to substantial shifts, even within an observation period of 20 years.

3.3 Productivity: technology or allocation?

Cross-entropy is a measure of how closely two distributions overlap. Recall, from Figure 1, that cross-entropy is minimized when the two distributions are are identical. This observation lends itself to a useful application to the decomposition of labour productivity. To see this, I first define a labour productivity index in Definition 3.

Definition 3 (Aggregate real productivity). *Real aggregate productivity growth $\Delta \log q_t$ equals GDP growth, minus labour input growth and an aggregate inflation index:*

$$\Delta \log q_t = \Delta \log Y_t - \Delta \sum_{i=1}^{N} \omega_{i,t} \log P_{i,t} - \Delta \log L_t. \tag{8}$$

A common choice for weights ω_i in the productivity literature is the two-period Törnqvist average, but I do not specify them here. Unlike before, productivity entails the use of two aggregate statistics; nominal output and labour, both of which are distributed among N industries. This means that the behaviour of the aggregate index $\Delta \log q_t$ hinges on the entropy of both nominal output and labour input. This leads to a convenient decomposition for labour productivity between : i)'technology' (how many more units are do industries produce per unit of labour, on average) and the allocation of ii) 'demand' (are expenditures for industries, that are weighted relatively more, relatively lower) and iii) labour (are labour input shares equal to industry weights). Proposition 2 outlines the components of this decomposition.

Proposition 2 (Aggregate real productivity decomposition). *Aggregate real productivity growth can be decomposed exactly between technology, plus the allocation of demand and labour:*

$$\Delta \log q_t = \underbrace{\Delta \sum_{i=1}^{N} \omega_{i,t} \log\left(\frac{Q_{i,t}}{L_{i,t}}\right)}_{\textit{Technology}} + \overbrace{\underbrace{\Delta D_{KL}(\mathbf{\Omega}_t \| \mathbf{Y}_t)}_{\textit{Demand}} - \underbrace{\Delta D_{KL}(\mathbf{\Omega}_t \| \mathbf{L}_t)}_{\textit{Labour}}}^{\textit{Allocation}}, \tag{9}$$

where $D_{KL}(x\|y)$ is the Kullback-Leibler (KL) divergence of y from x.

Proof. See [APPENDIX]. □

Proposition 2 predicts that demand allocation can contribute positively. This is reasonable in the sense that we want important industries, with larger weights, to cost less, and thus capture a smaller expenditure share. Labour allocation, however, contributes negatively; if the amount of labour inputs received by one industry does not match its importance in the weighting scheme, then there is a re-allocation opportunity that can decrease the KL divergence of labour from the weighting scheme towards zero. This decomposition is similar to that of Tang & Wang (2004), but benefits from the interpretation and additivity of log growth rates.

		$\Delta \log q_t$	Technology	Demand	Labour
France	1995-2005	1.64	1.58	0.14	-0.08
	2006-2017	0.64	0.73	-0.02	-0.08
	Slowdown	1.01	0.85	0.16	0.00
	Share	1.00	0.84	0.15	0.00
Germany	1995-2005	1.84	1.49	0.16	0.18
	2006-2017	0.87	0.96	0.02	-0.10
	Slowdown	0.97	0.54	0.15	0.28
	Share	1.00	0.55	0.15	0.29
Japan	1995-2005	1.75	1.48	0.12	0.15
	2006-2015	0.80	0.56	-0.00	0.25
	Slowdown	0.95	0.92	0.13	-0.11
	Share	1.00	0.98	0.13	-0.11
United Kingdom	1995-2005	2.18	1.90	0.18	0.11
	2006-2017	0.38	0.06	-0.02	0.34
	Slowdown	1.80	1.83	0.20	-0.23
	Share	1.00	1.02	0.11	-0.13
United States	1997-2005	2.45	2.58	0.22	-0.35
	2006-2017	0.88	1.02	-0.01	-0.13
	Slowdown	1.57	1.56	0.23	-0.22
	Share	1.00	0.99	0.15	-0.14

Table 3: **Accounting for allocation in the slowdown of labour productivity growth pre- and post-2005**; this table reports the sources of the labour productivity slowdown in five advanced economies, using the decomposition in Eq. 9. Weights are two-period Törnqvist averages of nominal output shares. Data from EU-KLEMS 2019.

Can the allocation terms be significant? I apply this decomposition to understand the infamous labour productivity slowdown, using data from EU KLEMS 2019 (Goldin et al. 2021). For the choice of indexation scheme, I use the conventional approach of two-period averages for the industries' nominal value added shares. Labour inputs are defined as number of hours worked. To summarise the problem, the first column reports real aggregate labour productivity growth (from a price index derived using appropriate weights) as an average for years pre- and post-2005, for France, Germany, Japan, the UK and the US. The slowdown is then defined as the difference between those two average growth rates; these range from around 1pp for the first three countries, to more than 1.5pp in the UK and US. How much of that slowdown can be explained by the allocation of demand, or labour?

A central finding is that the reallocation of demand appears to have worsened in all countries, explaining between 10 to 15% of the slowdown in all countries. The experience for labour reallocation, however, is more mixed; in Germany, labour reallocation worsened after 2005, and explains almost 30% of its labour productivity slowdown. However, labour reallocation does not appear to have changed significantly in France, and actually contributed positively to labour productivity in Japan, the UK and the US. In sum, this means that almost half of the slowdown in labour productivity for Germany can be explained by demand and labour reallocation. However, for the other countries the pure technology component – the ability of an average industry to produce more per hour worked – is the single explanation.

4 Conclusion

This paper demonstrates that accounting for variety matters when studying aggregate quantities. For example, past real output of a representative product from today is necessarily high, regardless of the inflation rate. Intuitively, this stems from the worsening rate of consumption of that product when going back in time. Similarly, real output today of a representative product from the past is low because the degree of representation worsens going forward in time. This bias is equal to the cross-entropy of GDP between the benchmark year and the rest of the series, which serves as an intuitive interpretation as the increasing loss of information from an index over time.

This information loss is significant in measuring GDP growth, even in highly-aggregated industry breakdowns. In US World KLEMS data, it amounts to a doubling in the real output index when indexing to current industry compositions of GDP, instead of those from seven decades earlier. This result is replicated in the shorter, yet more detailed, composition of CPI weights from 1998 to 2018. Consumer price data from Eurostat also suggests that, as a rule of thumb, real output growth rates extrapolated from GDP using aggregated inflation data are about 1-2pp higher compared to aggregated growth from observed real output data. Finally, a decomposition of labour productivity that incorporated the allocation of demand and labour inputs also provides some insight into the origin of the labour productivity slowdown for Germany.

The theme of the paper remains that variety, and accounting for it, matters for many topics in economics. There exist many more measures of variety that go beyond the simplest one considered here, namely entropy. This topic becomes more important now that many more granular datasets are becoming available for researchers to exploit, opening the frontier far beyond the study of singular, aggregate quantities.

References

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

Aghion, P. & Howitt, P. (1992), 'A model of growth through creative destruction', *Econometrica* **60**(2), 323–351.

Baqaee, D. & Burstein, A. (2021), Welfare and output with income effects and taste shocks, Working Paper 28754, National Bureau of Economic Research.

Bresnahan, T. F. & Gordon, R. J., eds (1997), *The Economics of New Goods*, University of Chicago Press.

Cover, T. M. & Thomas, J. A. (2005), *Entropy, Relative Entropy, and Mutual Information*, John Wiley & Sons, Ltd, chapter 2, pp. 13–55.

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

Eichhorn, W. (1978), *Systems of Functional Equations in the Theory of the Price Index*, Addison-Wesley Publishing Company, chapter 8, pp. 151–172.

Fisher, I. (1922), *The Making of Index Numbers*, Houghton-Mifflin.

Goldin, I., Koutroumpis, P., Lafond, F. & Winkler, J. (2021), 'Why is productivity slowing down?', *Working Paper* .

Gordon, R. J. (2009), Apparel prices 1914-93 and the hulten/bruegel paradox, *in* W. E. Diewert, J. S. Greenlees & C. R. Hulten, eds, 'Price Index Concepts and Measurement', University of Chicago Press, pp. 85–128.

Hulten, C. R. (1997), Comment on "do real-output and real-wage measures capture reality? the history of lighting suggests not", *in* Bresnahan & Gordon (1997), pp. 66–70.

Nordhaus, W. D. (1997), Do real-output and real-wage measures capture reality? the history of lighting suggests not, *in* Bresnahan & Gordon (1997), pp. 27–70.

Redding, S. J. & Weinstein, D. E. (2019), 'Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences', *The Quarterly Journal of Economics* **135**(1), 503–560.

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

Shannon, C. E. (1948), 'A Mathematical Theory of Communication', *The Bell System Technical Journal* **27**, 379 – 423.

Tang, J. & Wang, W. (2004), 'Sources of aggregate labour productivity growth in canada and the United States', *The Canadian Journal of Economics / Revue canadienne d'Economique* **37**(2), 421–444.

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

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