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*

February 20, 2023

Abstract

I introduce a general method to account for the distribution of underlying components (variety) in the growth of an aggregate quantity, using the notion of entropy. This accounting decomposition enables a number of insightful applications to 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 a decomposition of labour productivity growth by a technology component (how many more 'average' goods are produced per unit of labour?), and the allocation of labour (does the distribution of labour inputs converge to the distribution of outputs?) plus demand (does the distribution of expenditures diverge from the distribution of outputs?).

*__Acknowledgements__: I thank participants of the Economics Statistics Center of Excellence 2022 Conference, Rick Van der Ploeg, François Lafond, Ian Crawford, Erwin Diewert, J. Doyne Farmer and Martin McCarthy 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

Is Gross Domestic Product (GDP) a good measure for welfare? 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:

$$\text{Real output growth} = \text{Nominal output growth} - \text{Price Inflation.}$$

Nominal output is 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 hundreds of industries, millions of firms, and many more product types. 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, suggesting that 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 a good result in terms of minimising the influence of the cross-entropy bias.

This result applies to 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 two applications. I test the effect in a transparent manner, using industry price and output data from US KLEMS for the period 1947-2014. 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'.

Applying the notion of entropy to labour productivity is instructive for the wider applicability of entropy to economic questions. 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). Deaton (2010) famously re-draws the international poverty line to address differences in product variety across countries. 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. In this way, changing variety *solves* Hulten's paradox, instead of correcting it.

Information Theory concepts like entropy feature prominently in econometrics. Maasoumi (1993) offers a broader review, and explains their usefulness in studying the role of aggregation. This was an area of concern to Theil (1967) in his study of income inequality across population groups. In contrast, this paper focuses on distributional divergence over time, instead of a cross-sectional dimension.

Roadmap 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 extends it to study the aggregate productivity growth slowdown. Section 4 concludes.

2 Indexing variety

In this section, I develop a statistical accounting framework to aggregate nominal output – GDP – although it applies to any index that seeks to aggregate across a variety of types:

employment, exchange rates, monetary aggregates, etc. The role of variety emerges statistically as the *entropy* of the index, which theoretically evolves alongside incomes as long as preferences are not specifically Cobb-Douglas.

2.1 Entropy to measure variety

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 $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_N\}$, which I refer to as an *indexation scheme*. This is distinct from the distribution of industry nominal output shares

[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.

$y_i = Y_i/Y$, populating a $1 \times N$ vector $\mathbf{Y} = \{y_1, y_2, \ldots, y_N\}$, 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]

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

[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.

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, minimal 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 when applied to continuous time.

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.

2.2 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:

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

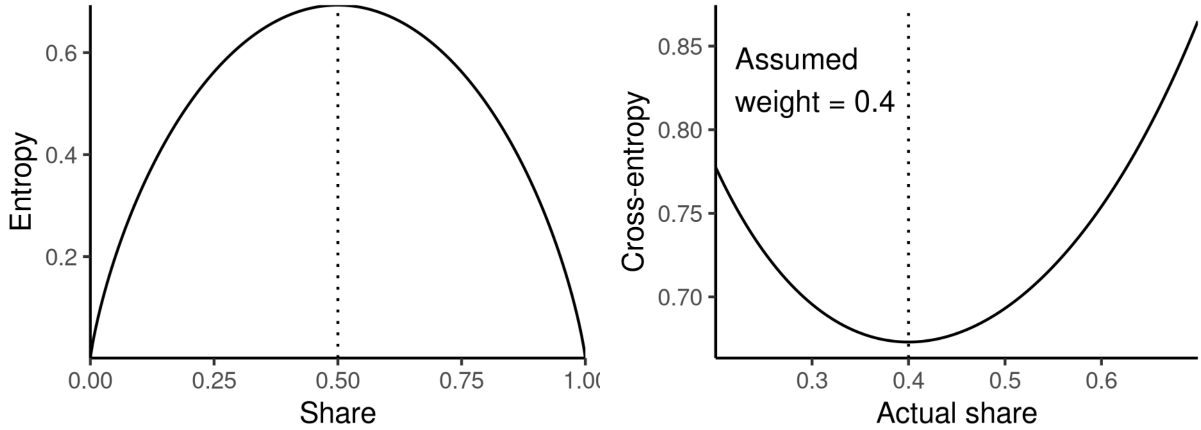

Figure 1: **Plotting entropy for two industries**: Total entropy (Left) is maximal at the dotted line, where the share of each industry is equal at 1/2. By fixing a weight (Right) at the dotted line, a change in the actual share increases cross-entropy.

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

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

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

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

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 $\mathbf{\Omega} \stackrel{d}{=} \mathbf{Y}$ 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

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

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, as per Aghion et al. (2019).

Comparing distributions Entropy is a metric for variety in one variable, but cross-entropy combines this measure for two distributions. The notion of 'accounting' for variety I propose is no more than a log-likelihood function for the distribution of GDP by industry against an assumed weighting scheme. In economics, these weights typically constitute a 'model'. Proposition 1 suggests that an accurate measure for aggregate real output requires a 'good' model when quantities are not directly observed by industry.

The right panel of Figure 1 demonstrates the extent to which variation in the true share of an industry, plotted along the horizontal axis, affects cross-entropy as measured by Eq. 4 when the assumed weight is fixed to 0.4. One issue with cross-entropy is that it is difficult to compare across different scenarios, because it is jointly sensitive to the performance of the model in addition to variety in the original data. Therefore, the *Kullback-Leibler* (KL) divergence is used as a common metric that standardises the divergence between two distributions (Kullback & Leibler 1951):

$$\begin{aligned} D_{KL}(\boldsymbol{\Omega} \| \mathbf{Y}) &= H(\boldsymbol{\Omega}, \mathbf{Y}) - H(\boldsymbol{\Omega}) \\ &= \sum_{i}^{N} \omega_i \log \frac{\omega_i}{y_i}, \end{aligned}$$

where $D_{KL}(x \| y)$ denotes the KL divergence of y from x. The desirable property of the KL divergence between two distributions is that it is zero when both distributions are identical. This is also seen from the minimised cross-entropy in Eq. 6. Further properties of the KL divergence measure are discussed by Cover & Thomas (2005)

2.3 Theoretical motivation

So far, I motivated the integration of entropy into macroeconomic indices by an intriguing accounting relationship. This comes with its own downside, since the result in Proposition 1 might strike the reader as tautological. Uneven growth rates must cause relative expenditure shares to diverge. However, theoretical assumptions also yield a variable level entropy, except under Cobb-Douglas aggregation. Entropy may therefore offer new insights into the nature of heterogeneous consumption in a demand system. I demonstrate this using a general framework.

Modelling consumption shares A parametric model can rationalise the share of consumption allocated to a given industry. Using such models, economists build 'cost of living' indices that aggregate industry prices to meet their benchmark 'standard of living', the corresponding aggregate real output index. Various assumptions can yield different parameterisations, and thus different weighting schemes by which to aggregate across industries. The main parameter that governs aggregation in these models is

$$\varepsilon_i = \frac{dY}{dQ_i}\frac{Q_i}{Y} \tag{8}$$

$$= \frac{Y_i}{Y} = y_i. \tag{9}$$

Parameter ε_i in Eq. 8 is defined as the *elasticity* of output Y with respect to quantity Q_i. Eq. 9 follows from assuming that Q_i is paid price P_i equivalent to its marginal output (the derivative in Eq. 8). This relationship is expected to hold at the system's equilibrium under perfect competition. Further extensions consider markups charged under imperfect competition, and other sources for 'wedges' (Baqaee & Farhi 2020).

Parameterisation using a utility function This elasticity features in a functional relationship between aggregate welfare and quantity inputs: a 'utility function' (or its dual with respect to prices, the 'cost function'). Writing this parameter explicitly is useful because it clearly delineates the realm of variability that these models offer under perfect competition. This is crucial for understanding the theoretical underpinning of entropy, which emerges from re-arranging Eq. 9:

$$\log Y = \log P_i + \log Q_i + \log\frac{1}{\varepsilon_i} \tag{10}$$

$$= \sum_{i=1}^{N}\omega_i \log P_i + \sum_{i=1}^{N}\omega_i \log Q_i + \sum_{i=1}^{N}\omega_i \log\frac{1}{\varepsilon_i}, \tag{11}$$

where $\sum_i \omega_i = 1$ as before. Eq. 10 demonstrates the earlier tautology: without measurement error, an index can be built from a single industry alone only if quantity growth is equal among all industries. In other words, aggregated output elasticities are residuals that only change when quantity growth is uneven among industries. The aggregation in Eq. 11 is desirable because it tracks 'aggregate' prices and quantities according to a scheme $\mathbf{\Omega}$, whose purpose can either be statistical – attenuating measurement error – or economic – tracing some aggregate 'welfare'.

When is entropy constant? It is important to note that different utility aggregators yield various predictions for what the elasticity should be. Therefore, changes in the output distribution can be interpreted in a variety of ways, depending on the assumptions on the utility function. Proposition 2 states a null result: what utility functions predict constant elasticities, and, consequently, a fixed output distribution?

Proposition 2 (Constant elasticity). *Aggregator $Y = f(Q_i)$ exhibits a constant elasticity ε_i with respect to quantity Q_i if, and only if, it holds functional form*

$$f(Q_i) = Q_i^{\varepsilon} c, \tag{12}$$

where c is constant with respect to Q_i.

Proof. See Appendix A.1.

□

Utility functions that follow the form of Eq. 12 are commonly referred to as 'Cobb-Douglas'. According to Proposition 2, Cobb-Douglas aggregators have constant elasticities that are dependent on a variety of goods, resulting in a constant entropy for the corresponding quantity index as income levels change. Preferences that generate constant expenditure shares as a function of income levels are known as *homothetic* preferences. This has significant implications, as it highlights that GDP alone is not a sufficient welfare metric unless preferences are Cobb-Douglas. Other utility functions require real output indices that are adapted to incorporate changes in preferences.

A prime example of such an adaptation is Baqaee & Burstein (2021). Their analysis considers the extent to which particular industries become more prevalent in consumption as incomes increase. Practical scenarios include healthcare and professional services, which grew as a share of GDP in advanced economies, in contrast to agricultural and certain industrial services, which generally shrank. Redding & Weinstein (2019) also introduce a demand system to assign an entropy-like residual, specifically for the Sato-Vartia index, some economic insight.

Does entropy increase over time? One reflection follows an established literature on the evolution of human wants as they progressively satisfy their needs (Maslow 1943). As spending on relatively niche outlets increases, and allocations turn more uniform, entropy invariably increases. More concretely, certain utility forms predict output elasticities as a function of input levels, allowing for some flexibility in assessing changes in output variety. Table 1 tabulates these elasticities for three popular functions used in microeconomic theory, with two inputs Q_1, Q_2. Section 3 offers empirical evidence for a broad trend in the historical entropy rate of the US economy.

One notable feature is that these utility functions, namely the Constant Elasticity of Substitution (CES) and the Translog, preserve the Cobb-Douglas as a special case. In both instances, output elasticities, and thus entropy, change if growth in utility components is uneven. This is readily seen in the third column of Table 1. In a CES scenario where Q_1 grows faster than Q_2, the expenditure share on Q_1 is predicted to decline if both components are complements ($\sigma < 0$), or increase if both are substitutes ($\sigma > 0$). This is due to the indirect effect on prices, thus breaking the tautology pointed out at the beginning of this section.

Table 1: **Demand elasticities for select utility functions**

Name	$f(Q_1, Q_2)$	ε_1
Cobb-Douglas	$AQ_1^{a_1}Q_2^{a_2}$	a_1
CES	$A\left[a_1 Q_1^{\sigma} + a_2 Q_2^{\sigma}\right]^{\frac{1}{\sigma}}$	$\frac{a_1 Q_1^{\sigma}}{a_1 Q_1^{\sigma} + a_2 Q_2^{\sigma}}$
Translog	$A + \sum_i a_i \log Q_i + \sum_{i,j} a_{i,j} \log Q_i \log Q_j$	$a_1 + \sum_i a_{1,i} \log Q_i$

Notes: Three candidate utility functions are tabulated, along the output elasticity ε_1 for one good Q_1. Constant coefficients are A, a generic multiplier, a_i, a technical coefficient, and σ, which yields an elasticity of substitution.

The Translog function obtains a similar result, where the signs of the second-order coefficients determine the direction for the expenditure share. Starting input levels will make a difference, since entropy will specifically increase(decrease) if expenditure shares grow more uniform(concentrated). It is possible to derive similar elasticities for other forms of utility.

2.4 Choosing weights

As per Corollary 1.1, changing indexation scheme $\mathbf{\Omega}_t$ can ameliorate, or worsen, the bias in real output in Proposition 1. However, one detail is that this corollary considers a static indexation scheme, built with information learnt in period t alone. Since quantity and price indices are defined by two-period changes, can we get around the bias by finding weights that minimise the change in cross-entropy, instead of its level?

Generally, yes. Table 2 categorises certain well-known indexation schemes, and demonstrates how cross-entropy enters the indexation formula. This variety in methods is due to the fact that some 'optimal' index should fulfill multiple criteria, which Eichhorn (1978) proves are impossible to satisfy simultaneously. Additionally, the desirability of certain indices depends on their interpretation within economic theory, or simple intuition.

Examples of weighting schemes A prominent choice for weighting scheme $\mathbf{\Omega}_t$ defined in continuous time are Divisia weights (Divisia 1925). Hulten (1973) demonstrates that, under the parameterisation in Eq. 9 plus an assumption on linear homogeneity, so that $\sum_i \varepsilon_{i,t} = 1$, these indices are path-independent. The Divisia index is conventionally approximated with two-period averages in expenditure shares for discrete time observations. Table 2 suggests that the bias due to cross-entropy under the Divisia approximation is small, specifically one half of the difference in the KL divergence of output shares in periods t and $t-1$. This term is large if fluctuations in output shares $y_{i,t}$ are large, but close to zero otherwise.

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. Specif-

Table 2: **Cross-entropy under different indexation schemes**

Name	$\omega_{i,t}$	$\Delta H(\mathbf{\Omega}_t, \mathbf{Y}_t)$
Divisia	$\frac{1}{2}(y_{i,t} + y_{i,t-1})$	$\frac{1}{2}[D_{KL}(\mathbf{Y}_{t-1} \Vert \mathbf{Y}_t) - D_{KL}(\mathbf{Y}_t \Vert \mathbf{Y}_{t-1})]$
Törnqvist	$\frac{1}{2}(y_{i,t1} + y_{i,t0})$	$\frac{1}{2}[D_{KL}(\mathbf{Y}_{t0} \Vert \mathbf{Y}_{t1}) - D_{KL}(\mathbf{Y}_{t1} \Vert \mathbf{Y}_{t0})]$
Sato-Vartia 2	$\frac{y_{i,t}-y_{i,t-1}}{\log y_{i,t}-\log y_{i,t-1}} \Big/ \sum_{i=1}^{N} \frac{y_{i,t}-y_{i,t-1}}{\log y_{i,t}-\log y_{i,t-1}}$	0
Chained	$y_{i,t-1}$	$D_{KL}(\mathbf{Y}_{t-1} \Vert \mathbf{Y}_t)$

Notes: The bias in real output due to cross-entropy in output shares, listed in the final column, varies by indexation scheme, named and expressed in the first two columns. $D_{KL}(x \Vert y)$ is the Kullback-Leibler (KL) divergence of y from x (Kullback & Leibler 1951, Cover & Thomas 2005).

ically, it assigns weights equal to the average output shares between two arbitrary periods – $t0$ and $t1$ in Table 2. Törnqvist weights coincide with the Divisa approximation when computed over rolling observations. 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 linearly the true weights change over time.

Whereas the Törnqvist index corresponds to Translog utility, the Sato-Vartia index corresponds to CES preferences (Sato 1976, Vartia 1976).[5] The Sato-Vartia index' own desirable property is that it adjusts for substitution between inputs via some constant elasticity, although at the price of violating Fisher's (1922) monotonicity axiom (Reinsdorf & Dorfman 1999). Redding & Weinstein (2019) further critique this index, demonstrating that it is problematic when preferences change over time. As far as accounting for variety is concerned, one feature that stands out in Table 2 is that the cross-entropy residual under Sato-Vartia indexation is precisely zero. The weights cancel the change in logarithmic output shares, and the remaining expression sums to zero.

Finally, I mention the chain index. The chain index, which updates weights to the previous period's output shares, is widely used in statistical agencies' CPI computations. The weights for consumption are updated annually and used to form monthly CPI estimates. This index provides a straightforward interpretation of price and quantity changes, as it is tied to a single representative basket from the recent past. This feature is particularly useful in high-frequency settings where there are persistent fluctuations, such as seasonal goods. In terms of accounting for variety, the cross-entropy term for a chained index reflects the simplest and most intuitive form for a change in entropy, which is the KL divergence of the current consumption basket from the previous basket.

[5] The specific version I consider is its second variant, Sato-Vartia 2.

2.5 Summary

This section demonstrates that 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.

Importantly, this bias remains present as long as preferences are not Cobb-Douglas. Therefore, this result suggests that adjusted real output indices are needed to account for varying preferences in order to obtain an accurate representation of economic activity. The main limitation in this argument is that it is derived assuming perfect competition. Recent studies investigate the role of markups and taxes that further distort the observed output distribution Baqaee & Farhi (2020). However, further adjustments for such wedges should manifest *in addition to* the cross-entropy term that relies on modelling output shares under perfect competition.

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 (1997) established the paradox in 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.

Solving the paradox 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.[6] He asserts that if true 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. In absolute terms, goods consumed in 1997 were likely scarce in colonial America – think of the entertainment a new

[6]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 products, 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.

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 3: **Yearly GDP growth with different industry indices: US, 1947-2014**

	$\Delta \log Y_t$	$\widetilde{P}_t$	$\widetilde{Q}_t$	$\Delta H(\mathbf{\Omega}_t, \mathbf{Y}_t)$
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
Chained	6.33	3.36	3.15	-0.18
Divisia	6.33	3.32	3.01	0.00

Notes: This table decomposes average nominal GDP growth from 1947 to 2014 between the average growth in price, volumes, and change in cross-entropy. Weights $\mathbf{\Omega}_t$ are labelled in the first column, and computed as per Table 2. All values denote log points. These observations aggregate across 65 industries, for which 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.

Data In order to capture the order of magnitude of this difference, I reproduce an aggregate GDP deflator from US World KLEMS data.[7] 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 respective 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 3 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. In addition, I include a Törnqvist index averaging 1947 and 2014, a chained index, and a rolling Törnqvist average to approximate a Divisia index, as outlined by Table 2.

Results The first column in Table 3 reports the average yearly nominal GDP growth rate, which is always constant. The second column reports the annual average price index, and the third the quantity index, constructed with the weighting scheme labelling each row. Using the annual industry-level nominal output shares, I retrieve the average cross-entropy

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

residuals corresponding to those schemes in the fourth column. Given that the decomposition in Proposition 1 is exact, the final three columns always add up to the nominal output growth rate reported in the first column.

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 in the second row, relative to 1947 in the first row. This difference, reported in the third row of the third column, is significant, in the order of 1pp per year. This translates to overstated real welfare improvements of about 50% after seven decades, if consumption patterns in 2014 are used as the benchmark.

The second, perhaps surprising outcome of this exercise is that aggregate deflators contribute *positively* to this gap. In the first and second rows of the second column, 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 CPI 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? The fourth row of Table 3 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, although slightly positive. This would suggest that industry shares evolved relatively smoothly during that period.

The fifth row reports the indices and cross-entropy for chained weights. As stated previously, this weighting scheme is perhaps most intuitive, since it benchmarks price and quantity growth to the most recent reporting period. The average rate of information loss between two subsequent years, which coincides with the KL divergence reported in Table 2, is relatively small at 0.18pp, almost 3% of average GDP growth.

Finally, the sixth row demonstrates that approximating a Divisia index yields a tiny cross-entropy term, although not exactly equal to zero. Therefore, this index is desirable for those who want to reduce the cross-entropy term to the greatest extent possible, short of using the Sato-Vartia 2 indexations scheme.

Appendix B.1 demonstrates that CPI weights from statistical agencies produce similar results, even though they include in the order of 200 product groups instead of coarse industry classes.

Sources of structural change It remains unclear whether these aggregate statistics conceal meaningful trends in the structure of the US economy over the last seventy years. Figure 2 plots the entropy rate of US GDP, $H(\mathbf{Y}_t)$, for years reported by US KLEMS. The underlying unit plotted on the vertical axis is the base of the natural logarithm, Euler's number, conventionally known as a 'nat' (Cover & Thomas 2005).[8] These industry data demonstrate a stunning reversal in the entropy rate, which has increase until the early 1980s, but since declined. This suggests that US industrial output has been most *uniform* in the early 1980s,

[8]Base two measures of entropy are known as 'bits', which are used to measure information in binary computing systems.

but turned relatively more concentrated in a few industries in recent years.

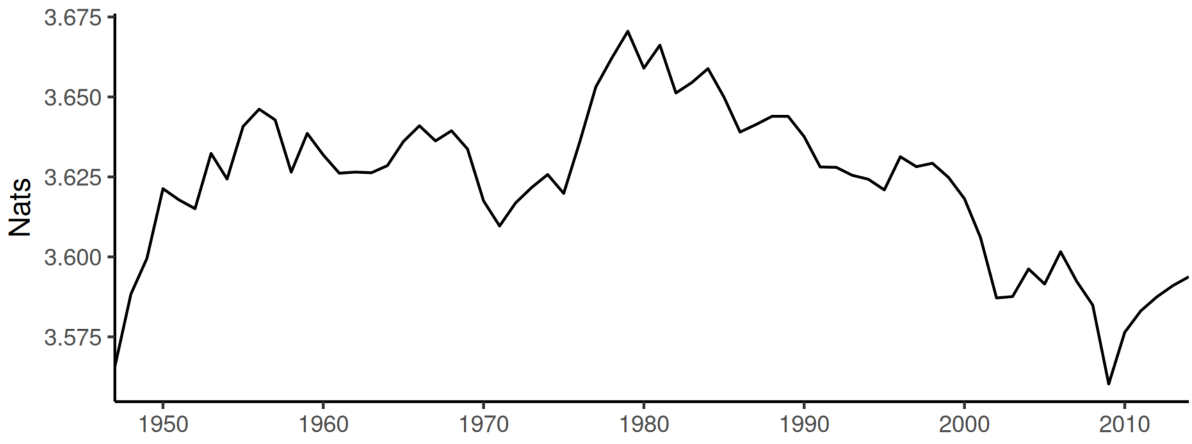

Figure 2: **Entropy of US GDP increases until 1980, but has declined since**: This figure plots the entropy of US GDP, measure in nats, from 1947 to 2014. Data are made available by World Klems.

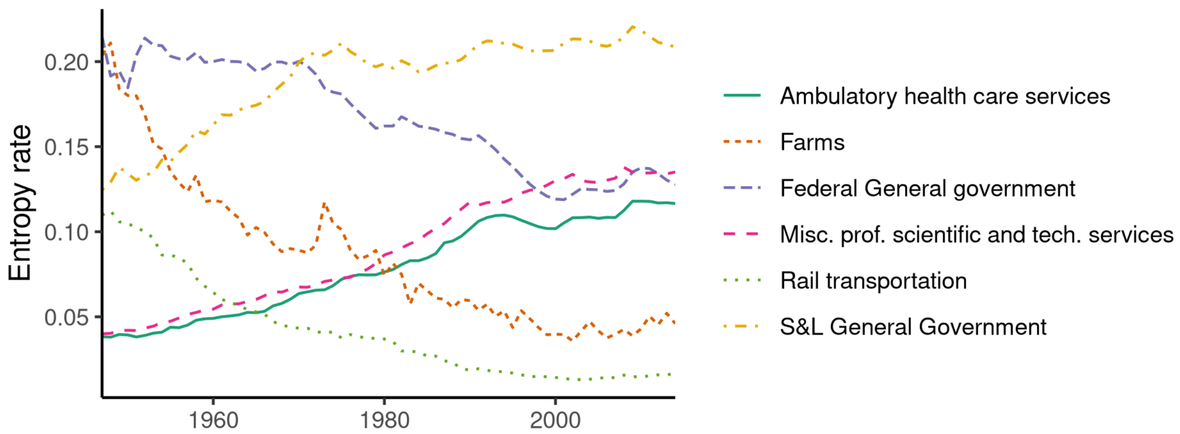

Figure 3: **Largest contributions to US entropy, 1947-2014**: This figure plots six industries that have the three largest positive and negative contributions to the entropy of US GDP. Data are made available by World Klems.

Figure 3 helps understand where most of the change in US entropy originated, by plotting the entropy rate

$$h_{i,t} = -y_{i,t} \log y_{i,t}$$

for select industries i. I plot this entropy rate for the three largest positive and negative contributions to the change in the total entropy rate observed over the whole sample by comparing the sums of entropy rates $\sum_{t=1947}^{2014} h_{i,t}$ between industries. The industries that emerge are well-documented in studies of structural change for 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, professional services, and S&L general government contributed most to an increasing entropy rate.

A promising avenue for further research constitutes aligning changes in the distribution and concentration of US industrial output with known drivers of structural change. One important phenomena is the emergence of IT industries, and service industries generally. A second driver is the emergence of global trade, which significantly reduced the importance of certain manufacturing industries. Another is cost disease (Baumol 1967), which likely affected agriculture.

3.2 Labour productivity: technology or allocation?

Cross-entropy is a measure of how closely two distributions overlap. Recall, from Figure 1, that cross-entropy is minimised 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 (Real productivity index)**.** *Real aggregate productivity growth $\Delta \log q_t$ equals GDP growth, minus labour input growth and an aggregate inflation index:*

$$\Delta \log q_t = \Delta \log Y_t - \Delta \sum_{i=1}^{N} \omega_{i,t} \log P_{i,t} - \Delta \log L_t. \tag{13}$$

A common choice for weights $\omega_{i,t}$ in the productivity literature is the rolling Törnqvist average, which coincides with the Divisia approximation expressed in Table 2. 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 an elegant decomposition of aggregate 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 3 outlines the components of this decomposition.

Proposition 3 (Aggregate real productivity decomposition)**.** *Aggregate real productivity growth can be decomposed exactly between technology, plus the allocation of demand and labour:*

$$\Delta \log q_t = \underbrace{\Delta \sum_{i=1}^{N} \omega_{i,t} \log\left(\frac{Q_{i,t}}{L_{i,t}}\right)}_{\textit{Technology}} + \overbrace{\underbrace{\Delta D_{KL}(\mathbf{\Omega}_t \| \mathbf{Y}_t)}_{\textit{Demand}} - \underbrace{\Delta D_{KL}(\mathbf{\Omega}_t \| \mathbf{L}_t)}_{\textit{Labour}}}^{\textit{Allocation}}. \tag{14}$$

Proof. See Appendix A.2. □

Cost disease Proposition 3 shares some similarities with the method presented by Tang & Wang (2004), but offers an improved interpretation and additivity of log growth rates. The decomposition predicts that demand allocation can contribute positively to the labour productivity growth index. This is reasonable in the sense that important industries, with larger weights, should cost less, and thus capture a smaller expenditure share. The demand allocation term constitutes an important link to Baumol's (1967) cost disease hypothesis, by which industries with low productivity growth capture a larger share of output. The precise version of the cost disease hypothesis that the demand allocation term captures will specifically depend on the choice of weights.

Labour misallocation While demand allocation contributes positively to the labour productivity growth index, labour allocation has a negative effect. This is because, if an industry receives a greater amount of labour inputs than its weight in the weighting scheme, it creates a reallocation opportunity by which relatively more labour input is sent to relatively important industries. In Proposition 3, this re-allocation can reduce the KL divergence of labour from the scheme towards zero. In contrast to Tang & Wang (2004), the decomposition does not focus exclusively on the productivity levels of industries. Rather, it measures reallocation according to the weighting scheme that is chosen, which does not have to reflect real output levels.

Explaining the slowdown in labour productivity growth Are the allocation terms significant for the infamous labour productivity slowdown? I use the conventional approach of two-period averages for the industries' nominal value added shares as the indexation scheme. 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. As Table 4 demonstrates, these allocation terms explain the labour productivity slowdown in varying proportions.

One of the key findings of the study is that the reallocation of demand appears to have worsened in all countries, explaining between 10 to 15% of the slowdown across the board. However, the experience with labour reallocation is more varied. In Germany, labour reallocation worsened after 2005, and explains almost 30% of the labour productivity slowdown. On the other hand, labour reallocation does not appear to have changed significantly in France, and has actually contributed positively to labour productivity in Japan, the UK, and the US. Taken together, these results indicate that demand and labour reallocation account for almost half of the slowdown in labour productivity for Germany. However, for the other countries, the pure technology component, which measures the average industry's ability to produce more per hour worked, is the single explanation.

4 Conclusion

In this paper, I highlight the importance of accounting for product variety in the analysis of aggregate quantities. I demonstrate that failing to consider the changing representation

Table 4: **Allocation in the slowdown of labour productivity growth**

		$\Delta \log q_t$	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

Notes: This table reports the sources of the labour productivity slowdown in five advanced economies, using the decomposition in Eq. 14. Weights are two-period Törnqvist averages of nominal output shares. Data from EU-KLEMS 2019.

of products over time can lead to biased estimates of real output. For instance, the real output growth experienced by a modern product may appear higher simply because the product was rarely consumed in the past, irrespective of its actual inflation rate. Similarly, the current real output growth of a representative product from the past may be underestimated due to its diminishing importance over time. This bias can be quantified by the cross-entropy of GDP between the benchmark year and the remaining period. I interpret this measure as the loss of information inherent in an index that accumulates over time. These findings highlight the need for careful consideration of product and industry variety in the study of aggregate quantities.

The loss of information due to accounting for variety has a significant impact on measuring GDP growth, even in fine-grained industry breakdowns. In the US World KLEMS data, for instance, it results in a doubling of the real output index level when using current industry compositions of GDP, compared to those from seven decades earlier. This finding is consistent with a more detailed analysis of CPI weights from 1998 to 2018, as well as with consumer price data from Eurostat, which suggests that extrapolated real output growth rates from aggregated inflation data are about 1-2pp higher than actual growth rates from observed real output data. Additionally, by incorporating the allocation of demand and labour inputs, a decomposition of labour productivity provides insight into the origins of the labour productivity slowdown in Germany since 2005.

The central argument of the paper is that variety plays a crucial role in analysing heterogeneous economic data, and this warrants careful consideration. Although the study focuses on a single application in the context of long-term economic growth, it suggests that accounting for variety could be relevant in other areas such as monetary aggregates and employment. There are several measures of variety that extend beyond the one examined in this paper, which could offer fresh perspectives. As researchers have access to increasingly granular datasets, this subject gains even greater significance, expanding the horizon beyond the analysis of singular aggregate quantities. The implications of this approach are broad, and further exploration could lead to new insights and innovative applications.

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.

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

Baqaee, D. R. & Farhi, E. (2020), 'Productivity and misallocation in general equilibrium', *Quarterly Journal of Economics* **135**(1), 105–163.

Baumol, W. J. (1967), 'Macroeconomics of unbalanced growth: The anatomy of urban crisis', *American Economic Review* **57**(3), 415–426.

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.

Deaton, A. (2010), 'Price indexes, inequality, and the measurement of world poverty', *The American Economic Review* **100**(1), 5–34.

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

Divisia, F. (1925), 'L'indice monétaire et la théorie de la monnaie', *Revue d'économie politique* **39**, 1121.

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.

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. (1973), 'Divisia index numbers', *Econometrica* **41**(6), 1017–1025.

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.

Kullback, S. & Leibler, R. A. (1951), 'On Information and Sufficiency', *The Annals of Mathematical Statistics* **22**(1), 79 – 86.

Maasoumi, E. (1993), 'A compendium to information theory in economics and econometrics', *Econometric reviews* **12**(2), 137–181.

Maslow, A. H. (1943), 'A theory of human motivation', *Psychological review* **50**(4), 370–396.

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.

Reinsdorf, M. B. & Dorfman, A. H. (1999), 'The Sato–Vartia index and the monotonicity axiom', *Journal of Econometrics* **90**(1), 45–61.

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.

Theil, H. (1967), *Economics and information theory*, Studies in mathematical and managerial economics ; v. 7, North-Holland ; Rand McNally, Amsterdam : Chicago.

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.

A Proofs

A.1 Proposition 2

Dropping subscript i from Eq. 8, elasticity

$$\varepsilon = \frac{dY}{dQ}\frac{Q}{Y} \tag{15}$$

is assumed constant. To derive the functional form of $Y = f(Q)$ that satisfies constant ε, re-arrange the expression, then integrate both sides:

$$\int f'(Q)dQ = \int \varepsilon \frac{f(Q)}{Q}dQ \tag{16}$$

$$= f(Q) + C = \varepsilon \int \frac{f(Q)}{Q}dQ \tag{17}$$

$$= \varepsilon f(Q)\log Q - \frac{\varepsilon^2}{2}f(Q)\log^2 Q + \frac{\varepsilon^3}{6}f(Q)\log^3 Q - \mathcal{O}(\log^4 Q) \tag{18}$$

$$= f(Q)\left[1 - \sum_{n=0}^{\infty}\frac{\varepsilon^n(-1)^n}{n!}(\log Q)^n\right] \tag{19}$$

$$= f(Q)\left[1 - \sum_{n=0}^{\infty}\frac{1}{n!}(\log Q^{-\varepsilon})^n\right] \tag{20}$$

$$= f(Q) - f(Q)\frac{1}{Q^{\varepsilon}}, \tag{21}$$

$$\Rightarrow f(Q) = Q^{\varepsilon}c, \tag{22}$$

where $C = -c$ is the constant of integration. Step 17 is possible because ε is assumed not to vary with Q. Step 18 follows from recursively integrating by parts and substituting $\varepsilon = Qf'(Q)/f(Q)$. The terms are collected in the sum of Step 19 after adding and subtracting one. This summation term is a Taylor expansion of $\exp(x)$ about zero, and thus simplifies to Step 21. Re-arranging yields the final result.

A.2 Proposition 3

Starting from productivity defined in Definition 3,

$$\Delta\log q_t = \Delta\log Y_t - \Delta\sum_{i=1}^{N}\omega_{i,t}\log P_{i,t} - \Delta\log L_t \tag{23}$$

$$= \Delta\sum_{i=1}^{N}\omega_{i,t}\log Q_{i,t} - \Delta\sum_{i=1}^{N}\omega_{i,t}\log L_{i,t} + \Delta\sum_{i=1}^{N}\omega_{i,t}\log\frac{Y_t}{Y_{i,t}} - \Delta\sum_{i=1}^{N}\omega_{i,t}\log\frac{L_t}{L_{i,t}} \tag{24}$$

$$= \Delta\sum_{i=1}^{N}\omega_{i,t}\log q_{i,t} + \Delta\sum_{i=1}^{N}\omega_{i,t}\log\frac{\omega_{i,t}Y_t}{Y_{i,t}} - \Delta\sum_{i=1}^{N}\omega_{i,t}\log\frac{\omega_{i,t}L_t}{L_{i,t}} \tag{25}$$

$$= \Delta\sum_{i=1}^{N}\omega_{i,t}\log q_{i,t} + \Delta D_{KL}(\mathbf{\Omega}_t\|\mathbf{Y}_t) - \Delta D_{KL}(\mathbf{\Omega}_t\|\mathbf{L}_t), \tag{26}$$

where $q_{i,t} = Q_{i,t}/L_{i,t}$, and $D_{KL}(x\|y)$ is the Kullback-Leibler (KL) divergence of y from x (Kullback & Leibler 1951).

B Additional results

B.1 Does the level of aggregation matter?

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.

BLS data 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.[9] 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 establishes 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 5: **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.

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

Results: US Table 5 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 3. 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 5 is the smaller inflation rate for the index based on 1998 consumption patterns.

Eurostat data 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.

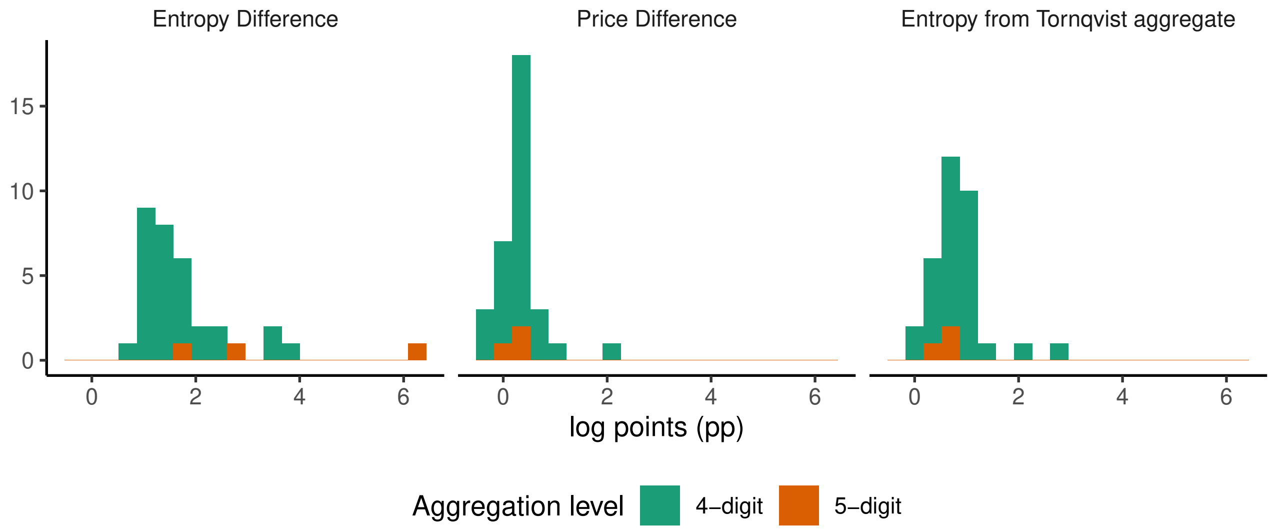

Figure 4: **Comparing price and entropy from first and last available weights in HICP data: 33 European countries, 1997-2020**: indexing aggregate price inflation to the start of the observation period yields an estimate for entropy (Left) that is around 1-2pp higher relative to an index composed of final-year weights. Overall inflation rates (Middle) are generally similar, if not slightly larger when indexing to final-year weights. A Törnqvist index (Right) averaging first- and final-year baskets lowers the average amount of cross-entropy in the index, but can still be in the order of 1pp.

Results: HICP By using weights from the first and final years available, Figure 4 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 fourth column in row four, of Table 5.

The middle chart in Figure 4 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 has an obvious impact on cross-entropy estimates, since 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 3 and in CPI weights in Table 5, may be special cases. In contrast, consumption patterns in other countries are subject to substantial shifts, even within an observation period of 20 years.