Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP01/2026

Proper Quantity Indexes for Coherent Productivity Comparisons

C.J. O'Donnell

Date: June 2026

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Proper Quantity Indexes for Coherent Productivity Comparisons

Christopher J. O'Donnell

School of Economics, University of Queensland

4 June, 2026

Abstract

This paper defines productivity as the ratio of an output quantity to an input quantity. Under this definition, a productivity index is the ratio of an output quantity index to an input quantity index. Coherent productivity comparisons across both time and space require quantity indexes that satisfy a small set of basic axioms. Quantity indexes satisfying these axioms are referred to as *proper* indexes. Such indexes can be constructed using a wide range of aggregator functions and weighting systems.

There is no uniquely correct proper index. Coherent productivity comparisons require only that empirical aggregation choices preserve the axioms necessary for consistent quantity comparisons. Many index-number formulas widely used in applied work are designed for pairwise or one-dimensional comparisons (over time *or* across space) and generally fail to satisfy these axioms when comparisons extend across multiple periods or units. Because multi-period and multi-unit comparisons are ubiquitous in applied economics, the framework described in this paper has implications well beyond productivity analysis.

The paper clarifies the conceptual foundations of productivity comparisons, reviews the principal classes of productivity measures used in practice, and explains why statistical agencies and applied researchers continue to rely on measures that are not proper.

Keywords: measurement theory; multilateral comparisons; multitemporal comparisons.
JEL classification: C43; D24

1 Introduction

Measures of productivity provide a compact, quantity-based summary of economic performance. They are widely used to compare the performance of businesses, industries, regions, and countries, and to assess changes over time and across space. Such uses impose stringent requirements on how productivity is measured. If productivity statistics are to support meaningful performance comparisons, the underlying measures must faithfully represent the relationships among the observations being compared. This paper treats that requirement as a problem of measurement rather than one of economic modelling or statistical estimation. The remainder of the paper adopts this measurement-theoretic perspective.

Although the focus of this paper is productivity measurement, the central problem addressed here is one of quantity measurement more broadly. Comparisons across time and space in applied economics routinely rely on aggregated measures of output and input quantities. The validity of such comparisons depends on the properties of the underlying quantity measures. The measurement framework described in this paper therefore has implications not only for productivity analysis, but for any empirical setting in which quantities are aggregated and compared.

The paper is organised as follows. Section 2 clarifies the quantity-based definition of productivity adopted in the paper and places it in the context of alternative definitions that are not generally quantity-based. Section 3 sets out the measurement-theoretic framework that underpins quantity-based productivity concepts, including the axioms required for coherent quantity comparisons across time and space. Section 4 defines the concept of a proper quantity index and describes the principal classes of such indexes, including additive, multiplicative, primal, dual, and benefit-of-the-doubt formulations. Section 5 contrasts these with quantity indexes that are widely used in empirical work but do not, in general, satisfy the axioms required for coherent multilateral and multitemporal comparisons.[1] Section 6 clarifies the relationship between endpoint index-number comparisons and growth-rate representations of productivity change, and identifies the conditions under which chained growth rates yield coherent multitemporal comparisons. Section 7 examines how productivity is measured in practice, including the use of non-proper indexes and Malmquist-type measures, and illustrates the practical implications of these choices with a simple numerical example. Section 8 concludes.

[1] In this paper, the term "multilateral" is used to refer to comparisons across spatial units, while the term "multitemporal" is used to refer to comparisons across time periods. Comparisons involving both spatial and temporal variation are therefore described as "multilateral and multitemporal" comparisons. This terminology differs somewhat from the broader usage in parts of the index-number literature, where "multilateral" is often used to refer to any comparison involving more than two observations.

2 Definitions of Productivity

The term 'productivity' is used in the economics literature and in political discourse to refer to a range of related but distinct concepts. These concepts differ in their purpose, their interpretation, and the types of comparisons they are intended to support. This section reviews common definitions of productivity used in economics and places them in their historical and analytical context.

This paper defines productivity as the ratio of an output quantity to an input quantity. This definition underpins the index-number tradition in productivity analysis and is stated explicitly or implicitly in influential contributions to that literature (e.g., Christensen, 1975; Diewert, 1976; Caves et al., 1982a; Balk, 1998; OECD, 2001). It is also consistent with definitions found in the broader economics literature (e.g., Chambers and Pope, 1996; Syverson, 2011). Measures of output per unit of input include labour productivity (output divided by labour input), multifactor productivity (output divided by an aggregate of some, but not all, inputs), and total factor productivity (output divided by an aggregate of all factors of production). Productivity so defined is a quantity ratio that can be computed for each observation. Comparisons of productivity across observations are made using productivity indexes. Because productivity is defined as the ratio of an output quantity to an input quantity, productivity indexes are ratios of output quantity indexes to input quantity indexes. The need to construct output and input quantity indexes—and therefore productivity indexes—that permit coherent comparisons across time and space motivates the measurement-theoretic approach adopted in this paper.

One influential alternative definition of productivity arises in the growth-accounting framework of Solow (1957). Solow specifies a production relationship in which output growth is driven by changes in inputs and by technical change, where the term *technical change* is used as shorthand for any shift in the production function over time. Under his maintained assumptions of constant returns to scale and Hicks-neutral technical change, this unobserved technical change component can be interpreted as output divided by an aggregate input—that is, as productivity in the quantity-based sense adopted in this paper. Interpreting technical change as productivity change has since become widespread in the empirical economics literature (e.g., Diewert, 1992; Barro, 1999; Basu et al., 2001) Under Solow's additional assumptions of competitive factor markets and successful cost minimisation, the aggregate input can be evaluated using observed cost shares, allowing productivity to be computed as a residual—what has become known as the *Solow residual.* This interpretation of the residual as productivity is conditional: it relies on restrictive assumptions that effectively rule out important sources of productivity change, including technical efficiency and economies of scale and substitution.

Over time, this conditional interpretation has given way to a more expansive usage in which residuals from a wide range of econometric models are labelled as productivity, even when the restrictive assumptions underlying the Solow framework do not hold. For example, Olley and Pakes (1996) write that "productivity is defined here, as elsewhere, as the residual from a relationship between deflated sales and inputs" (p. 1264); they then proceed to treat this residual as an unobserved state variable in the firm's decision problem, in a setting where the production function does not exhibit constant returns to scale. Other examples can be found in a large empirical literature in applied industrial organisation, macroeconomics, and plant-level productivity analysis (e.g., Petrin and Levinsohn, 2012; Syverson, 2011; Ackerberg et al., 2015; Harris and Moffat, 2015). In these papers, the residual no longer admits a structural interpretation as a ratio of output quantity to input quantity; instead, it is best interpreted as unobserved statistical noise.[2]

A closely related definition is developed by Jorgenson and Griliches (1967), who begin by defining the rate of growth of total factor productivity as the difference between the rate of growth of real product and the rate of growth of real factor input (Jorgenson and Griliches, 1967, p. 250). Because rates of growth correspond to logarithmic changes in index numbers, this definition is consistent with the productivity index framework developed in this paper. Jorgenson and Griliches also provide an economic interpretation of productivity growth; this interpretation is based on competitive equilibrium conditions and a production function that exhibits constant returns to scale. Within this framework, shifts in the production function are again interpreted as changes in productivity. As in the Solow framework, this interpretation relies on assumptions that rule out technical inefficiency and economies of scale or substitution. Once these assumptions are relaxed, shifts in the production function need not coincide with changes in a properly defined output–input ratio.

Relaxing the assumption of full technical efficiency leads naturally to a broader characterisation of productivity change. For example, Färe et al. (1994) allow producers to operate below the production frontier and "define productivity growth as the product of efficiency change and technical change" (p. 72). Under this definition, productivity growth reflects both shifts in the production frontier and movements toward or away from that frontier. While this represents an important generalisation of the Solow and Jorgenson–Griliches frameworks, it continues to abstract from productivity changes associated with economies of scale and substitution.

Relaxing the assumption of constant returns to scale leads to yet another characterisation of productivity change. For example, Hranaiova and Stefanou (2002) argue that "productivity growth … can be expressed as a sum of technical change and scale effects" (p. 79). In this case, productivity growth reflects both shifts in the production frontier and changes in economies of scale. By allowing for variable returns to scale, this

[2] In econometric models, statistical noise arises from functional-form errors, measurement errors, omitted-variable errors, and included-variable errors, the individual contributions of which are not generally identifiable without strong assumptions.

approach departs from the Solow and Jorgenson–Griliches frameworks. However, it continues to exclude productivity changes arising from technical inefficiency and economies of substitution. As a result, the associated measures cannot in general be interpreted as measuring changes in the output–input ratio.

All of these alternative definitions share a common feature: they define productivity in terms of a particular explanation of productivity change. In the Solow and Jorgenson–Griliches frameworks, productivity is identified with technical change; in the framework of Olley and Pakes (1996), it is identified as an unobserved state variable in the firm's decision problem; in the framework of Färe et al. (1994), it is identified with technical change and technical efficiency change; and in the framework of Hranaiova and Stefanou (2002), it is identified with technical change and scale effects. In each case, productivity is defined not as a quantity-based measure of performance, but as a model-dependent explanation of changes in that performance. This conflation of measurement and explanation is understandable, but it has two important consequences. First, once productivity is defined in terms of a particular set of drivers, other potential sources of productivity change are excluded by definition rather than assessed empirically. As a result, alternative definitions of productivity proliferate as different assumptions are relaxed, even though the underlying object of interest—output per unit of input—remains unchanged. Second, it becomes natural to describe productivity growth as a source of output growth (e.g., Ball et al., 2016; USDA, 2023). Interpreted literally, this claim is difficult to reconcile with the quantity-based definition adopted in this paper. If productivity is defined as the ratio of an output quantity to an input quantity, then productivity growth cannot, in a logical sense, constitute an independent source of output growth. Claims about the role of productivity in economic growth are therefore meaningful only in special cases where productivity change coincides with one or more explanatory components embedded in a specific production model. The most common case is the identification of productivity change with technical change. Under that interpretation, statements about the role of productivity in economic growth are more precisely understood as statements about the role of technical progress.

The alternative definitions discussed so far all arise within quantity-based production frameworks. Other measures that arise within such frameworks include Malmquist–Luenberger productivity indicators, which are constructed using directional distance functions (e.g., Chambers et al., 1996; Chung et al., 1997; Boussemart et al., 2003; Epure et al., 2011). Although these measures are quantity-based, they do not generally evaluate changes in output–input ratios. Rather, they measure changes in directional distance-function values determined by a chosen direction of adjustment toward the production frontier. Because productivity is defined in this paper as the ratio of an output quantity to an input quantity, such difference-based indicators do not constitute measures of productivity change under the definition adopted here.

A more fundamental departure occurs when productivity is defined in terms of revenues or costs rather than quantities. In this case, the object of measurement itself changes—from quantities to values. An example of a revenue-based definition is provided by Garcia-Marin and Voigtländer (2019), who calculate a measure of total factor productivity as "the residual between total revenues and the estimated contribution of production factors (labor, capital, and material inputs)" (p. 1787). Similar measures can be found in Klette (1999), Foster et al. (2008) and Hsieh and Klenow (2009). Only in special cases—most notably when output prices are common across observations—can revenues be interpreted as aggregate output quantities, in which case revenue per unit of input coincides numerically with output per unit of input. In general, however, revenue-based measures combine variations in quantities with variations in prices and therefore cannot be interpreted as quantity-based measures of productivity.

A further class of measures arises when measures of output are replaced by measures of *real value added* (e.g., Diewert and Fox, 2018; Adamopoulos et al., 2022). Real value added is obtained by deflating nominal value added by a price index. Nominal value added is defined as the value of total output minus the cost of intermediate inputs. The classification of inputs as intermediate or primary depends on the level at which production is defined.[3] Regardless of whether inputs are classified as primary or intermediate, two practical issues arise. First, deflating a nominal value by a price index does not, in general, produce a valid measure of quantity change; the resulting series will not generally satisfy the axioms required for coherent quantity comparisons across time or space. Second, even if real value added were to provide a coherent measure of quantity change, it is constructed by subtracting the deflated value of intermediate inputs from the deflated value of outputs rather than by aggregating outputs alone. Ratios based on real value added therefore do not coincide with ratios of output quantity to input quantity and are better interpreted as measures of real value added per unit of input rather than as quantity-based measures of productivity.

More broadly, it is useful to distinguish clearly between ratios of quantities and ratios involving values. Productivity, as defined in this paper, is a ratio of quantities. Ratios involving values are standard objects of economic analysis, but they are conventionally given different names. Ratios of values are commonly referred to as measures of profitability or rates of return (such as return on equity or return to the dollar), while ratios of values to quantities are often referred to as average revenues and average costs. Cost- and revenue-based constructs therefore do not represent alternative ways of measuring

[3]At the economy-wide level, production is typically represented as transforming primary inputs—labour and capital—into outputs, with energy, materials, and services treated as intermediate inputs because they are themselves produced within the economy and used in subsequent stages of production. At narrower levels of analysis, such as a local government area or establishment, these same inputs enter production as externally supplied inputs and therefore function as primary inputs from the local perspective.

productivity, but instead measure different economic objects.

This discussion highlights a fundamental distinction between defining productivity as an object of measurement and explaining observed changes in that object using particular economic models. When productivity is defined directly as a ratio of output and input quantities, the central problem becomes one of measurement: how to construct coherent, internally consistent measures of productivity levels and productivity change across space and time. This problem is inherently measurement-theoretic. It requires explicit attention to index-number properties, aggregation, comparability, and invariance, independent of any particular behavioural or structural interpretation. The following section sets out the measurement-theoretic framework adopted in this paper. It takes the quantity-based definition of productivity as given and treats the economic explanation of productivity change as a conceptually distinct and logically subsequent step.

3 Measurement Theory

Taking the definition of productivity from the previous section as given raises a fundamental question: how should heterogeneous quantities be aggregated so that productivity comparisons are meaningful? Answering this question requires a theory of measurement. Measurement theory is a branch of mathematics concerned with assigning numbers to objects in a way that preserves empirically observed relationships among those objects. In the present context, the objects of interest are vectors of quantities, and the purpose of measurement is to support coherent comparisons across both time and space.

Once quantity indexes are understood as numerical representations of empirical relationships among quantity vectors, it becomes necessary to specify the properties such representations must satisfy. Properties that are regarded as self-evidently necessary for coherent numerical comparisons are referred to here as axioms. In this sense, axioms are not empirical hypotheses or behavioural assumptions, but normative requirements that a numerical representation must satisfy in order to preserve the structure of the phenomena being measured. Not all commonly used index-number formulas satisfy the axioms required for general comparisons across both time and space.

To fix ideas and avoid unnecessary duplication, this section formulates the axioms using a generic quantity vector q, which may represent either outputs or inputs. Let q_{it} denote a vector of quantities associated with firm i in period t, where the term *firm* is used generically to refer to a production unit. A quantity index $QI(\cdot,\cdot)$ is a mapping from ordered pairs of quantity vectors into the set of nonnegative real numbers, intended to represent relative quantities in a way that preserves empirically observed relationships. The axioms below can be found in O'Donnell (2018).

Axiom 1 (Weak monotonicity). Weak monotonicity requires that a quantity index be nondecreasing in each component of the comparison bundle, holding the reference bundle fixed. If one quantity vector dominates another componentwise, then the quantity index must represent the former as weakly larger than the latter relative to any common benchmark. This axiom ensures that quantity index formulas respect dominance relations in the data and rule out representations in which strictly larger quantity bundles are mapped to smaller numerical values. Mathematically, weak monotonicity requires that

$$q_{rl} \geq q_{it} \;\Rightarrow\; QI(q_{ks}, q_{rl}) \geq QI(q_{ks}, q_{it}). \tag{1}$$

Axiom 2 (Homogeneity). Homogeneity requires that scaling the comparison bundle by a positive factor scales the quantity index by the same factor. This axiom ensures that quantity indexes respond appropriately to uniform expansions or contractions of the comparison bundle. Mathematically, homogeneity requires that for any $\lambda > 0$,

$$QI(q_{ks}, \lambda q_{it}) = \lambda\, QI(q_{ks}, q_{it}). \tag{2}$$

Axiom 3 (Scale Invariance). Scale invariance requires that a common proportional rescaling of the two quantity vectors being compared leaves the quantity index unchanged. This axiom ensures that quantity comparisons do not depend on the overall scale of measurement when all commodities are rescaled by the same factor. Without this property, quantity comparisons could depend on arbitrary scaling conventions rather than on underlying quantities. Mathematically, scale invariance requires that for any $\lambda > 0$,

$$QI(\lambda q_{ks}, \lambda q_{it}) = QI(q_{ks}, q_{it}). \tag{3}$$

Axiom 4 (Proportionality). Proportionality requires that when one quantity vector is a scalar multiple of another, the quantity index that compares them equals that scalar. This axiom anchors quantity measurement to transparent benchmark cases and ensures consistency between quantity index comparisons and direct proportional changes in observed quantities. Mathematically, proportionality requires that for any $\lambda > 0$,

$$QI(q_{ks}, \lambda q_{ks}) = \lambda. \tag{4}$$

Axiom 5 (Time–space reversal). Time–space reversal requires that reversing the roles of the two quantity vectors being compared inverts the value of the quantity index.

This axiom ensures that quantity comparisons are independent of arbitrary choices about reference observations and supports consistent comparisons across both time and space. Mathematically, time–space reversal requires that

$$QI(q_{ks}, q_{it}) = \frac{1}{QI(q_{it}, q_{ks})}. \tag{5}$$

Axiom 6 (Transitivity). Transitivity requires internal consistency of quantity comparisons across multiple observations. If one quantity vector is compared to a second, and the second to a third, then the implied comparison between the first and third must coincide with the direct comparison. This axiom is essential for coherent multilateral and multitemporal comparisons and for the construction of rankings that are free of logical contradiction. Mathematically, transitivity requires that

$$QI(q_{ks}, q_{rl})\, QI(q_{rl}, q_{it}) = QI(q_{ks}, q_{it}). \tag{6}$$

The axioms set out above are requirements on a *quantity index*: a numerical representation of relationships among quantity vectors alone. A different tradition, associated with Fisher (1922), evaluates index number formulas using a collection of 'tests' that involve both prices and quantities. As shown by Swamy (1965) and Eichhorn (1976), Fisher's tests are mutually inconsistent: no single index-number formula can satisfy all of them simultaneously. From a measurement-theoretic perspective, however, the difficulty lies in the strength of the requirements imposed by Fisher's tests. In particular, Fisher's criteria combine conditions that are necessary for coherent quantity comparisons with additional restrictions related to prices and values. The axioms used in this paper are deliberately weaker: they are restricted to properties that are necessary for coherent numerical representation of quantity relationships and do not impose additional price-based or value-based constraints. Precisely because they are weaker, these axioms can be satisfied simultaneously.

A further source of debate concerns transitivity (and the closely related circularity property, implied by transitivity and proportionality). Some authors treat transitivity as a discretionary *desideratum* in multiple comparison problems, to be traded off against other properties of index numbers such as 'characteristicity' (i.e., the requirement that the weights used in a comparison should be representative of the observations being compared; e.g., Drechsler, 1973), or against economically plausible properties of production technologies such as variable returns to scale and non-neutral technical change (e.g., Färe and Grosskopf, 1996, p. 91). In a representational measurement framework, transitivity is not a dispensable auxiliary criterion. It is the requirement that the set of pairwise comparisons generated by an index be internally consistent: comparisons must not depend on the path taken through intermediate observations. When the purpose of measurement is to support coherent comparisons across both time and space, path-independence is part of what it means for a numerical representation to preserve the relationships among the phenomena being compared.

4 Proper Quantity Indexes

Sections 2 and 3 defined productivity in quantity terms and set out the basic axioms required for coherent quantity comparisons across time and space. A *proper* quantity index is defined here to be any index-number formula that satisfies those axioms, with no further restrictions imposed. The term *proper* was introduced in O'Donnell (2016) to describe quantity indexes satisfying an equivalent axiomatic system; the formulation used here adopts a streamlined set of axioms and updated terminology, but the concept is unchanged.

This definition of a proper quantity index admits a wide class of index formulas. A proper quantity index comparing quantities for observation (i, t) with those for observation (k, s) is defined as

$$QI(q_{ks}, q_{it}) \;=\; \frac{Q(q_{it})}{Q(q_{ks})}, \tag{7}$$

where $Q(\cdot)$ is any nonnegative, nondecreasing, linearly homogeneous, scalar-valued aggregator function. Additive, multiplicative, primal, dual, and benefit-of-the-doubt (BoD) quantity indexes all fall within this definition. Differences among them arise from alternative choices of aggregator functions and weights, reflecting analytical convenience, data availability, and normative or interpretative judgements about how heterogeneous quantities should be combined.

The purpose of this section is not to identify a single preferred index, but to describe the main classes of proper quantity indexes and clarify how they differ. All of the indexes discussed below satisfy the axioms set out in Section 3; their differences are best understood as differences in aggregation choices.

4.1 Additive Quantity Indexes

Additive quantity indexes aggregate heterogeneous quantities using linear aggregator functions:

$$Q(q_{it}) \;\propto\; \sum_{n=1}^{N} w_n\, q_{nit}, \tag{8}$$

where $q_{it} = (q_{1it}, \ldots, q_{Nit})'$ denotes the vector of quantities for observation (i, t) and $w = (w_1, \ldots, w_N)'$ is a vector of non-negative, observation-invariant weights. The term *additive* refers to a structural property of the resulting index formula: when the reference quantity vector is held fixed, the index number generated for the sum of two comparison quantity vectors equals the sum of the index numbers generated for each comparison vector individually. A formal statement of this property is given in equation (1) of Aczel and Eichhorn (1974); a closely related dual property for the reference quantity vectors—expressed in terms of inverse index numbers—is given in their equation (2).

Different choices of weights generate different additive quantity indexes. When average market prices are used as weights, the resulting indexes take the form of the Lowe quantity indexes defined by Hill (2008).[4] A prominent example is the Geary–Khamis (GK) quantity index used in the International Comparison Program (ICP): it can be interpreted as a Lowe index that uses average international market prices as weights (Hill, 2008, p. 18). Applications of Lowe indexes in empirical productivity analysis include O'Donnell (2012) and Pan and Walden (2015). Alternatively, weights can be generated mechanically using statistical methods. For example, principal components analysis (PCA) can be used to choose weights so that the resulting index has maximal variance among all linear combinations of the component quantities, without requiring information on prices, value shares, or production technologies. Empirical applications of PCA-weighted indexes include Nagar and Basu (2002) and Feroze and Chauhan (2010).

4.2 Multiplicative Quantity Indexes

Multiplicative quantity indexes aggregate quantities using double-log aggregator functions:

$$Q(q_{it}) \propto \prod_{n=1}^{N} q_{nit}^{w_n}, \tag{9}$$

where the non-negative weights sum to one. Again, the term *multiplicative* refers to a structural property of the associated index: when the reference quantity vector is held fixed, the index number generated for the Hadamard (element-wise) product of two comparison quantity vectors equals the product of the index numbers generated for each comparison vector individually. A formal statement of this property, and of the corresponding dual property for the reference quantity vectors, is given in O'Donnell (2018, fn. 9, p. 96).

Once more, different choices of weights generate different multiplicative quantity indexes. When average value (i.e., revenue or cost) shares are used as weights, the resulting

[4]The literature does not always distinguish additive index forms according to how weights are constructed, and the label "Lowe index" is sometimes applied to additive indexes more generally. In this paper, the term *additive quantity index* denotes the class of indexes defined by an additive aggregation structure, while *Lowe quantity index* refers specifically to an additive index that employs average market prices as weights.

indexes are sometimes referred to as geometric Young indexes.[5] Alternatively, weights may be generated mechanically—for example, within a regression framework—yielding multiplicative indexes whose weights need not reflect economic value shares or an underlying production structure (see, e.g., O'Donnell, 2018 for discussion).

4.3 Primal Quantity Indexes

Primal output and input quantity indexes aggregate quantities using output and input distance functions. The term *primal* reflects the fact that distance functions are commonly referred to as primal representations of production technologies. In the case of outputs, a period-and-environment-specific output distance function gives the reciprocal of the largest factor by which a given output vector can be proportionally expanded when using a given input vector in a given period and production environment.[6] Formally, the reciprocal of the largest factor by which the output vector q can be scaled up when using the input vector x in period t and production environment z is defined as

$$D_O^t(x, q, z) = \inf\{\rho > 0 : x \text{ can produce } q/\rho \text{ in period } t \text{ in environment } z\}. \tag{10}$$

By construction, this function is nonnegative and linearly homogeneous in outputs. If outputs are strongly disposable, then it is also nondecreasing in outputs. In this case, it can be used as an aggregator function for constructing a proper output quantity index. There are two natural ways forward. One possibility is to follow O'Donnell (2018, p. 97) and define

$$Q(q_{it}) \;\propto\; D_O^{\bar{t}}(\bar{x}, q_{it}, \bar{z}), \tag{11}$$

where $\bar{t}$ is a fixed reference period and $\bar{x}$ and $\bar{z}$ are fixed reference vectors of inputs and environmental variables. If the output set is homothetic and technical change is Hicks neutral, then the distance function is separable in outputs and the output index is therefore invariant to the choice of $\bar{t}$, $\bar{x}$ and $\bar{z}$.

If the distance function is continuously differentiable, then a second possibility is to compute the associated normalised shadow output prices,

$$p_n^t(x, q, z) = \frac{\partial D_O^t(x, q, z)}{\partial q_n}, \qquad n = 1, \ldots, N, \tag{12}$$

and then use a linear aggregator function with average shadow prices as weights. The

[5] Usage of the term "geometric Young" varies across the index-number literature and is not always tied to a specific choice of weights. Here, *multiplicative quantity index* denotes the class of indexes having a multiplicative aggregation structure, while *geometric Young index* refers specifically to the case in which comparison-invariant average value shares are used as weights.

[6] An *environmental variable* is one that is physically involved in the production process but never under the control of the firm manager (e.g., rainfall in farming; see O'Donnell, 2016, p. 329).

rationale for using a linear function follows from Euler's Theorem for homogeneous functions, which implies that

$$D_O^t(x, q, z) = \sum_{n=1}^{N} p_n^t(x, q, z)\, q_n. \tag{13}$$

The rationale for averaging the shadow prices is that $p_n^t(x, q, z)$ generally varies with x, q, z, and t, as the notation makes explicit. In the additive construction considered here, however, the weights in the linear aggregator function must be observation invariant. Averaging the shadow prices across observations produces a fixed vector of weights consistent with this requirement. An index constructed in this way shares the same additive structure as a Lowe quantity index, but uses average shadow prices rather than average market prices as weights. Because ratios of normalised shadow output prices correspond to marginal rates of transformation, the resulting index can also be interpreted as an additive quantity index that aggregates outputs using average marginal rates of transformation as weights.

An entirely analogous construction applies on the input side. A primal input quantity index can be constructed using a period-and-environment-specific input distance function, which gives the largest factor by which a given input vector can be proportionally scaled down while producing a given output vector in a given period and production environment. When the input distance function is continuously differentiable, the associated normalised shadow input prices can be used to construct additive input quantity indexes using averages of those shadow prices as weights. Because ratios of normalised shadow input prices correspond to marginal rates of technical substitution, the resulting indexes can equivalently be viewed as additive input quantity indexes that aggregate inputs using average marginal rates of technical substitution as weights. Once again, these indexes share the additive structure of a Lowe input quantity index, but use average shadow prices rather than average market prices as weights.

4.4 Dual Quantity Indexes

Dual output and input quantity indexes aggregate quantities using cost and revenue functions. The term *dual* reflects the fact that cost and revenue functions are commonly referred to as dual representations of production technologies, in contrast to the primal representations provided by distance functions.

In the case of outputs, a period-and-environment-specific cost function gives the minimum cost of producing a given output vector in a given period and production environment. Formally, if a firm is a price taker in input markets and faces the input price vector

$w \geq 0$, then the minimum cost of producing q in period t in production environment z is

$$C^t(w, q, z) = \min_x \{w'x : x \text{ can produce } q \text{ in period } t \text{ in environment } z\}. \qquad (14)$$

Under standard free-disposability assumptions, this function is nonnegative and nondecreasing in outputs. If the period-and-environment-specific output distance function is homogeneous of degree $-r$ in inputs, then $C^t(w, q, z)$ is homogeneous of degree $1/r$ in outputs. This homogeneity property allows the r-th power of the cost function to be used as an aggregator function for constructing a proper output quantity index.

Again, there are two natural ways forward. The first is to follow O'Donnell (2018, p. 98) and define the aggregator function

$$Q(q_{it}) \;\propto\; C^{\bar{t}}(\bar{w}, q_{it}, \bar{z})^{\,r}, \qquad (15)$$

where $\bar{t}$ is a fixed reference period and $\bar{w}$ and $\bar{z}$ are fixed reference vectors of input prices and environmental variables. If the input set is homothetic and technical change is Hicks neutral, then the cost function is separable in outputs. Under these regularity conditions, the resulting output quantity index is invariant to the choice of $\bar{t}$, $\bar{w}$, and $\bar{z}$.

If the cost function is continuously differentiable, a second possibility is to compute the marginal costs,

$$MC_n^t(w, q, z) = \frac{\partial C^t(w, q, z)}{\partial q_n}, \quad n = 1, \ldots, N, \qquad (16)$$

and then use a linear aggregator function with average marginal costs as weights. Again, the rationale for using a linear function is Euler's Theorem, which in this case implies that

$$C^t(w, q, z) = r \sum_{n=1}^{N} MC_n^t(w, q, z)\, q_n. \qquad (17)$$

As in the primal case, averaging is required because marginal costs generally vary with w, q, z, and t; averaging of marginal costs over observations produces a fixed set of weights. An index constructed in this way can be viewed as an additive quantity index that uses average marginal costs as weights.

An entirely analogous construction applies on the input side. If the period-and-environment-specific output distance function is homogeneous of degree $-r$ in inputs, then—under standard free-disposability assumptions—the revenue function for a price-taking firm is nonnegative, nondecreasing in inputs, and homogeneous of degree r in inputs. Dual input quantity indexes can be constructed using such functions. If the revenue function is continuously differentiable, averages of marginal revenues—derivatives of the revenue function with respect to inputs—can be used as weights to construct additive input quantity indexes.

4.5 Benefit-of-the-Doubt (BoD) Quantity Indexes

Benefit-of-the-Doubt (BoD) quantity indexes provide a mechanical approach to aggregating heterogeneous quantities that differs fundamentally from the approaches discussed above. Rather than requiring a fixed vector of weights—be they prices, value shares, or marginal effects—BoD indexes allow the weights to vary across observations in a tightly controlled way.

In the output case, a BoD aggregator function assigns weights to each observation so as to make its aggregate output as large as possible, subject to the constraint that the same weights, if applied to any other observation in the dataset, would not yield a larger aggregate output. Formally, for each observation (i, t), the weights are chosen to solve the linear program

$$\max_{a_{it}} \{a'_{it} q_{it} : a_{it} \geq 0,\ a'_{it} q_{hs} \leq 1 \text{ for all observations } (h, s)\}, \tag{18}$$

where a_{it} is a vector of non-negative weights and q_{it} denotes the corresponding vector of quantities. The maximised value of the objective function is the aggregate output for observation (i, t) and gives that observation the *benefit of the doubt*.

This construction has two important features. First, BoD indexes are entirely mechanical: in particular, they do not rely on any economic functions or assumptions about markets or firm behaviour. Although BoD indexes are often computed using data envelopment analysis (DEA) software, this is purely a matter of computational convenience. The linear programs used in the BoD approach resemble those used in DEA, but the interpretation is index-number–theoretic rather than production-theoretic, and the resulting weights are not estimates of economic objects.

Second, BoD indexes relax the requirement that aggregation weights be invariant across observations. Unlike additive, multiplicative, primal, or dual indexes—which impose a common set of weights reflecting prices, shares, or marginal effects—BoD indexes allow the weights to differ from one observation to the next. This flexibility is carefully managed: the admissible weights are restricted by linear constraints that preserve the axioms required for proper quantity indexes.

An entirely analogous construction applies on the input side, where the weights are chosen to minimise aggregate inputs subject to the constraint that no observation can appear to use fewer inputs than any other when evaluated using the same weights. BoD output and input quantity indexes are therefore best viewed as providing upper-envelope measures of aggregate outputs and lower-envelope measures of aggregate inputs. They are particularly useful in settings where no information is available on prices, value shares, or production technologies, or where the analyst wishes to avoid imposing any such structure. Formal properties and further discussion are available in the benefit-of-the-doubt literature (e.g., Cherchye et al., 2007; van Puyenbroeck and Rogge, 2017; Rogge, 2018).

5 Non-Proper Quantity Indexes

The previous section showed that a broad class of proper quantity indexes is available for productivity analysis. Nevertheless, much applied work relies on alternative index-number formulas that do not satisfy the axioms required for coherent quantity comparisons in multilateral or multitemporal settings. This section describes the main classes of such *non-proper* quantity indexes, outlines how they are constructed, and indicates the axioms they violate. The purpose is descriptive rather than evaluative; the practical consequences of using non-proper indexes are examined in Section 7.

5.1 Binary Quantity Indexes

Binary quantity indexes are designed for situations in which the objective is to make a single comparison between two observations. They therefore typically use comparison-specific weights that depend on the particular pair of observations being compared. In the index-number literature, these weights are typically constructed from observed prices or value shares. For example, the binary Fisher (BF) and binary Törnqvist (BT) indexes that compare quantities for observation (i,t) with those for observation (k,s) are defined as:

$$QI^{BF}(q_{ks}, q_{it}) = \left(\frac{\sum_{n=1}^{N} p_{nks}\, q_{nit}}{\sum_{n=1}^{N} p_{nks}\, q_{nks}} \; \frac{\sum_{n=1}^{N} p_{nit}\, q_{nit}}{\sum_{n=1}^{N} p_{nit}\, q_{nks}} \right)^{1/2}, \tag{19}$$

and

$$QI^{BT}(q_{ks}, q_{it}) = \prod_{n=1}^{N} \left(\frac{q_{nit}}{q_{nks}} \right)^{\frac{1}{2}(s_{nit}+s_{nks})}, \tag{20}$$

where $p_{it} = (p_{1it}, \ldots, p_{Nit})'$ denotes a vector of prices and $s_{it} = (s_{1it}, \ldots, s_{Nit})'$ is the associated vector of value shares. The BF index satisfies many of the binary tests proposed by Fisher (1922), and for this reason it has been frequently recommended and often used in empirical productivity analysis (e.g., Diewert, 1992; Ray and Mukherjee, 1996; Kuosmanen and Sipiläinen, 2009). The BT index has also been widely recommended and used in empirical work (e.g., O'Mahony and Timmer, 2009; Salim and Islam, 2010; Feenstra et al., 2015). It is commonly motivated by results due to Caves et al. (1982b), who show that, under restrictive regularity conditions, the BT index is equivalent to a Malmquist index.

Malmquist output and input quantity indexes are constructed using distance functions rather than observed prices and value shares. On the output side, let $D_O^t(x, q)$ denote the period-specific output distance function, defined as the reciprocal of the largest factor by which the output vector q can be radially expanded using the input vector x in period t. This is the conventional period-specific output distance function used in the Malmquist literature, which does not explicitly condition on environmental variables such as those introduced earlier in Sections 4.3 and 4.4. A binary Malmquist output quantity index comparing observation (i, t) with observation (k, s) can be written in the form

$$QI^{OM}(q_{ks}, q_{it}; x_{ks}, x_{it}) = \left[\frac{D_O^s(x_{ks}, q_{it})}{D_O^s(x_{ks}, q_{ks})} \frac{D_O^t(x_{it}, q_{it})}{D_O^t(x_{it}, q_{ks})}\right]^{1/2}, \tag{21}$$

where x_{it} denotes the input vector for observation (i, t). If only one firm is involved in the comparison (i.e., $i = k$), the firm subscripts can be suppressed and the two ratios in (21) correspond to the middle (ratio) terms in Diewert (1992, Eqs. (118)–(119)). A corresponding Malmquist input quantity index is defined analogously using a period-specific input distance function.

Importantly, binary indexes are not transitive, except in restrictive special cases where they reduce to proper indexes: the BF index reduces to a Lowe index if and only if prices are constant across observations; the BT index reduces to a geometric Young index if and only if value shares are constant across observations; and the Malmquist output quantity index reduces to a primal index if and only if output sets are homothetic and technical change is Hicks neutral. It follows that the Fisher, Törnqvist and Malmquist quantity indexes cannot support internally consistent multiple comparisons unless their respective restrictive conditions hold.

5.2 Chained Quantity Indexes

Chained quantity indexes are mainly used in time-series settings, where comparisons are often made sequentially between adjacent periods and then linked together to form longer-run comparisons. For example, the chained Fisher (CF) and chained Törnqvist (CT) indexes that compare quantities for observation (i, t) with those for observation (i, s) are defined as:

$$QI^{CF}(q_{is}, q_{it}) = \prod_{r=s}^{t-1} QI^{BF}(q_{ir}, q_{i,r+1}), \tag{22}$$

and

$$QI^{CT}(q_{is}, q_{it}) = \prod_{r=s}^{t-1} QI^{BT}(q_{ir}, q_{i,r+1}), \tag{23}$$

where $QI^{BF}(\cdot, \cdot)$ and $QI^{BT}(\cdot, \cdot)$ are the binary Fisher and Törnqvist indexes defined above. These indexes can be motivated operationally by the convenience of constructing multitemporal comparisons by linking adjacent-period (binary) comparisons, each of which depends only on data observed in two consecutive periods. More fundamentally, however, chain-linking is usually justified by the view that economic structures evolve over time, so that no single set of weights is appropriate for long-term comparisons (e.g., Landefeld et al., 1995; OECD, 2001, Ch. 7). Unfortunately, chained indexes are path dependent and do not, in general, satisfy the proportionality axiom in Section 3.[7] Accordingly, chained indexes do not, in general, provide a coherent basis for multiple comparisons across time (and, *a fortiori*, across both time and space). They can support such comparisons only in restrictive special cases in which the underlying binary indexes reduce to proper indexes. In those cases, the chained indexes are also proper.

5.3 Multilateral Quantity Indexes

Multilateral quantity indexes are commonly used in cross-section settings, where many observations are compared within a single period and a complete matrix of pairwise comparisons is required. For example, the Eltetö–Köves–Szulc (EKS) and Caves–Christensen–Diewert (CCD) indexes that compare quantities for observation (i, t) with those for observation (k, t) are defined as:

$$QI^{EKS}(q_{kt}, q_{it}) = \left[\prod_{r=1}^{I_t} QI^{BF}(q_{kt}, q_{rt}) QI^{BF}(q_{rt}, q_{it})\right]^{1/I_t}, \tag{24}$$

and

$$QI^{CCD}(q_{kt}, q_{it}) = \left[\prod_{r=1}^{I_t} QI^{BT}(q_{kt}, q_{rt}) QI^{BT}(q_{rt}, q_{it})\right]^{1/I_t}, \tag{25}$$

where I_t denotes the number of observations available in period t. These constructions are widely recommended and used in international comparisons of quantities and productivity (e.g., Caves et al., 1982b; Fox, 2003; Sickles and Zelenyuk, 2019; McCormack et al., 2020). They arise from attempts to impose transitivity on multilateral comparisons constructed

[7] Nor do they generally satisfy the transitivity axiom. Instead, they satisfy a temporal linking identity of the form $QI^C(q_{is}, q_{it}) = QI^C(q_{is}, q_{iu}) QI^C(q_{iu}, q_{it})$ for $s < u < t$. This identity is sometimes described as "transitivity", but it is path dependent: the chained comparison generally depends on the choice of intermediate observations.

from binary indexes. However, multilateral indexes formed as geometric averages of binary comparisons generally fail the proportionality axiom in Section 3. Consequently, they do not, in general, provide a coherent representation of quantity differences across multiple observations. Such consistency arises only in restrictive special cases in which the underlying binary indexes reduce to proper indexes; in those cases, the multilateral index is also proper.

5.4 Implicit Quantity Indexes

Implicit quantity indexes are commonly constructed by deflating value aggregates using price indexes. This approach is standard practice in national accounts and empirical productivity analysis when direct quantity data are unavailable or difficult to measure (e.g., Eurostat, 2020; Australian Bureau of Statistics, 2021). Unfortunately, implicit quantity indexes generally fail the axioms required for proper quantity measurement. It is not the value comparisons that are problematic. Value aggregates are linear functions of quantities, and value indexes provide coherent measures of value change. The difficulty arises only when value indexes are deflated by price indexes in an attempt to obtain a measure of quantity change. Such deflation does not generally yield a proper quantity index. O'Donnell (2018, Sec. 3.4) describes restrictive special cases under which primal output and input quantity indexes are equal to revenue and cost indexes divided by dual output and input price indexes. Except in such special cases, dividing a value index by a price index does not yield a proper quantity index and therefore does not support internally consistent multiple comparisons.

6 Rates of Productivity Growth

Measures of productivity change are measures of output quantity change divided by measures of input quantity change. The preceding sections discuss methods for measuring quantity change (and therefore productivity change) across time and space using endpoint index numbers. A large empirical and theoretical literature instead measures productivity change over time using growth rates. This emphasis reflects the fact that many research and policy questions are framed in terms of changes in productive performance over time, rather than comparisons across space. Much of this literature derives growth-rate measures by differentiating production, cost, revenue, or distance functions. The resulting expressions take the form of Divisia measures of output growth relative to input growth. From a measurement-theoretic perspective, growth-rate measures are simply an alternative way of making intertemporal comparisons. This section clarifies the relationship between growth-rate representations and index-number comparisons and identifies the conditions under which chained growth rates yield coherent endpoint comparisons.

6.1 Growth Rates Versus Endpoint Index Numbers

In a discrete-time setting, productivity comparisons between two observations can be constructed directly using endpoint index numbers, or indirectly by chaining a sequence of period-to-period growth factors. Chaining yields a coherent endpoint index-number comparison only if the period-to-period growth factors are themselves consistent with a single underlying comparison system. This requirement can be stated formally as follows. Let $\{g_{t,t+1}\}_{t=1}^{T-1}$ denote a sequence of measured one-period growth factors. These growth factors yield coherent endpoint comparisons if and only if there exists a positive scalar sequence $\{I_t\}_{t=1}^{T}$ such that $g_{t,t+1} = I_{t+1}/I_t$ for all $t = 1, \ldots, T-1$. In this case, chained comparisons telescope, so that $\prod_{k=s}^{t-1} g_{k,k+1} = I_t/I_s$ and the implied endpoint comparison is independent of the choice of intermediate periods used for chaining.

In index-number theory, this requirement is known as *transitivity.* Transitivity ensures that comparisons between any two observations are internally consistent and do not depend on the choice of intermediate observations. In continuous-time settings, growth-rate representations are typically expressed as time derivatives, while endpoint index-number comparisons correspond to integrating these local growth rates along a path over an interval. The continuous-time analogue of transitivity is *path independence*: the endpoint comparison implied by integrating local growth rates must depend only on the endpoints and not on the particular path taken between them. The issue of path independence is often discussed explicitly in the productivity-growth literature (e.g., Feng and Serletis, 2010).

6.2 Construction Using Distance Functions

It is common to construct measures of productivity growth by differentiating the logarithm of a production or distance function (e.g., Star, 1974, p. 124; Skevas, 2023, Sect. 3.2). Endpoint comparisons can then be obtained by integrating the resulting local growth rates over time. To fix ideas, suppose the output distance function defined in (10) is nondecreasing in outputs, nonincreasing in inputs, and continuously differentiable along a path $\{x(t), q(t), z(t)\}$. Define $\dot{v}(t) \equiv \frac{d}{dt} \ln v(t)$. The Appendix shows that the total differential of $\ln D_O^t(x(t), q(t), z(t))$ can be rearranged to yield the Divisia-type local productivity-growth expression

$$\sum_{n=1}^{N} \phi_n(t)\, \dot{q}_n(t) - \sum_{m=1}^{M} \gamma_m(t)\, \dot{x}_m(t) = TP_O(t) + EC_O(t) + OTEC(t), \tag{26}$$

where $\phi_n(t) \equiv \partial \ln D_O^t / \partial \ln q_n \geq 0$ is a shadow revenue share, $\gamma_m(t) \equiv -\partial \ln D_O^t / \partial \ln x_m \geq 0$ is an output elasticity satisfying $\sum_{m=1}^{M} \gamma_m(t) = 1$ under constant returns to scale, $TP_O(t) \equiv -\partial \ln D_O^t / \partial t$ is a measure of output-oriented technical progress, $EC_O(t) \equiv$

$-\sum_\ell(\partial \ln D_O^t/\partial \ln z_\ell)\dot{z}_\ell(t)$ measures output-oriented environmental change, and $OTEC(t) \equiv \frac{d}{dt}\ln D_O^t(x(t), q(t), z(t))$ measures output-oriented technical efficiency change. The corresponding endpoint comparison between periods s and t is given by the exponential of the integrated local growth rate:

$$\frac{I_t}{I_s} \equiv \exp\left\{\int_s^t \left[\sum_{n=1}^{N} \phi_n(\tau)\,\dot{q}_n(\tau) - \sum_{m=1}^{M} \gamma_m(\tau)\,\dot{x}_m(\tau)\right] d\tau\right\}. \tag{27}$$

In general, this endpoint comparison is path dependent: unless the weights $\{\phi_n(t)\}$ and $\{\gamma_m(t)\}$ are consistent with a single underlying comparison system, the value of I_t/I_s can vary with the path $\{x(\tau), q(\tau), z(\tau)\}_{\tau=s}^{t}$. A strong sufficient condition for path independence is that the weights $\{\phi_n(t)\}$ and $\{\gamma_m(t)\}$ are globally invariant (i.e., constant). In the growth accounting literature it is common to assume that output sets are homothetic, technical change is Hicks neutral, and production functions exhibit constant returns to scale. These regularity conditions are sufficient to ensure the output elasticities $\{\gamma_m(t)\}$ do not vary with outputs, and the shadow-revenue shares $\{\phi_n(t)\}$ do not vary with inputs, environmental variables, or time.[8] However, these conditions are not generally sufficient to ensure that the weights are globally invariant. If they *are* globally invariant and sum to one, then (27) is equivalent to the ratio of a multiplicative output index to a multiplicative input index. Since multiplicative indexes are proper indexes, this ratio supports internally consistent measurement of productivity growth.

6.3 Construction Using Cost and Revenue Functions

Another approach to constructing measures of productivity growth involves differentiating the logarithm of a cost or revenue function (e.g., Schreyer, 2010). As in the distance-function setting, this yields Divisia-type local growth-rate representations, and endpoint comparisons can be obtained by integrating the resulting local growth rates over time. To illustrate, suppose the cost function defined in (14) is nondecreasing in outputs, nondecreasing in input prices, and continuously differentiable along a path $\{w(t), q(t), z(t)\}$. The Appendix shows that the total differential of $\ln C^t(w(t), q(t), z(t))$ can be rearranged

[8] This paper is not concerned with *explaining* productivity change. However, it is worth noting that if the weights *are* globally invariant, then the regularity conditions discussed here also rule out a separate scale-and-mix efficiency change term on the right-hand side of (26), and in any coherent endpoint decomposition of productivity change.

to yield the Divisia-type local productivity-growth expression

$$\sum_{n=1}^{N} \beta_n(t)\,\dot{q}_n(t) - \sum_{m=1}^{M} s_m(t)\,\dot{x}_m(t) = TP_C(t) + EC_C(t) + CEC(t) + \sum_{m=1}^{M} \big[s_m(t) - \alpha_m(t)\big]\,\dot{w}_m(t), \tag{28}$$

where $\beta_n(t) \equiv \partial \ln C^t / \partial \ln q_n \geq 0$ is the n-th cost elasticity satisfying $\sum_{n=1}^{N} \beta_n(t) = 1$ under constant returns to scale, $s_m(t) \equiv w_m(t)x_m(t)/C(t) \geq 0$ is the m-th observed cost share, $\alpha_m(t) \equiv \partial \ln C^t / \partial \ln w_m \geq 0$ is the m-th cost-minimising cost share, $TP_C(t) \equiv -\partial \ln C^t / \partial t$ measures cost-oriented technical progress, $EC_C(t) \equiv -\sum_{j=1}^{J} (\partial \ln C^t / \partial \ln z_j) \dot{z}_j(t)$ measures cost-oriented environmental change, $CEC(t) \equiv \frac{d}{dt} \ln CE(t)$ measures cost-efficiency change, and the final term is a price wedge reflecting differences between observed and cost-minimising cost shares. The corresponding endpoint comparison between periods s and t is again given by the exponential of the integrated local growth rate:

$$\frac{I_t}{I_s} \equiv \exp\left\{ \int_s^t \left[\sum_{n=1}^{N} \beta_n(\tau)\,\dot{q}_n(\tau) - \sum_{m=1}^{M} s_m(\tau)\,\dot{x}_m(\tau) \right] d\tau \right\}. \tag{29}$$

Once more, the endpoint comparison in (29) is generally path dependent: unless the weights used to construct the local growth rates are consistent with a single underlying comparison system, the value of I_t/I_s will vary with the path $\{w(\tau), x(\tau), q(\tau), z(\tau)\}_{\tau=s}^{t}$. As in the distance-function case, a strong sufficient condition for path independence is global invariance of the relevant weights. Under standard regularity conditions—homothetic input sets, Hicks-neutral technical change, and constant returns to scale—the cost elasticities $\beta_n(t)$ do not vary with input prices, and the cost-minimising cost shares $\alpha_m(t)$ do not vary with outputs, environmental variables, or time.[9] However, these conditions are not generally sufficient to ensure global invariance. If they are globally invariant and satisfy the adding-up conditions required for proportionality, then (29) reduces to the ratio of multiplicative output and input indexes.

An analogous construction applies on the revenue side. In that setting, a productivity-growth expression is obtained by differentiating the logarithm of the revenue function and combining it with the derivative of observed revenue. As before, endpoint comparisons are obtained by integrating the resulting local growth rates over time. Coherent endpoint comparisons require path independence, and this requirement will be met if the relevant weights are globally invariant.

[9] If the weights are globally invariant, then the regularity conditions discussed here, together with the standard assumption of cost-minimising behaviour, rule out a separate scale-and-allocative-efficiency-change component in (28) and in any coherent endpoint decomposition of productivity change. Under cost-minimising behaviour, the cost-efficiency-change term and the price-wedge term also vanish.

7 Productivity Measures in Empirical Practice

The preceding sections have argued that productivity measures intended to support multiple comparisons across time and space should be based on proper quantity indexes. Yet much empirical work, including official statistics, continues to rely on measures that do not satisfy the axioms required for proper quantity measurement. This section explains why this divergence between measurement theory and empirical practice matters, and why it persists.

7.1 Proper Productivity Measurement

Taking the quantity-based definition of productivity as given, productivity change must be measured as the ratio of an output quantity index to an input quantity index. This approach underlies official productivity statistics and the construction of productivity accounts in national and international statistical systems (e.g., OECD, 2001; Australian Bureau of Statistics, 2024).

Following O'Donnell (2016, p. 332), and throughout this paper, a productivity index is said to be *proper* if and only if it can be written as the ratio of a proper output quantity index to a proper input quantity index. When this condition holds, the resulting productivity measure provides a coherent basis for internally consistent comparisons of productive performance across time and space.

Importantly, adopting proper productivity indexes typically does not require information beyond the quantities, prices, and value shares already used to construct non-proper alternatives. In most applications, the binding constraints are not data limitations but inherited conventions and default software implementations that treat long-run multiple-comparison problems as sequences of unrelated binary comparisons.

In principle, the requirement that productivity be measured as the ratio of a proper output quantity index to a proper input quantity index does not dictate a single formula. It identifies a class of admissible constructions. In empirical work, proper indexes may be implemented directly using additive or multiplicative formulas with comparison-invariant weights, using primal or dual representations based on distance, cost, or revenue functions, or using benefit-of-the-doubt (BoD) formulations in which weights are allowed to vary across observations in a tightly controlled manner. The choice among these approaches is largely pragmatic and depends on the available data, the dimensionality of outputs and inputs, and the intended use of the resulting comparisons.

7.2 Malmquist Productivity Indexes

A widely used alternative to proper productivity measurement is the construction of Malmquist productivity indexes (e.g., Färe et al., 1994; Coelli and Rao, 2005; Zelenyuk and Zhao, 2025). Unlike the measures considered above, the Malmquist productivity index is not constructed directly as the ratio of an output quantity index to an input quantity index. For example, the standard output-oriented Malmquist productivity index comparing observation (i, t) with observation (k, s) may be written as

$$MPI(q_{ks}, q_{it}; x_{ks}, x_{it}) = \left[\frac{D_O^s(x_{it}, q_{it})}{D_O^s(x_{ks}, q_{ks})} \frac{D_O^t(x_{it}, q_{it})}{D_O^t(x_{ks}, q_{ks})} \right]^{1/2}, \tag{30}$$

where $D_O^t(\cdot)$ denotes the period-specific output distance function defined in Section 5.1. Malmquist productivity indexes are widely used in empirical work because they admit decompositions into measures of technical change and technical efficiency change. In the approach of Färe et al. (1994), this decomposition effectively serves as the definition of productivity change. However, the resulting measure does not, in general, coincide with productivity change defined as output quantity change divided by input quantity change. For example, Grifell-Tatjé and Lovell (1995) demonstrate that, in the presence of variable returns to scale, Malmquist productivity indexes provide systematically biased measures of productivity change. Even when constructed under constant returns to scale, Malmquist productivity indexes are binary indexes that are not generally transitive.

Several other Malmquist productivity indexes have been proposed in the literature (e.g., Färe et al., 1992; Färe and Zelenyuk, 2021). Some are input-oriented, and some are alternative output-oriented forms that are algebraically equivalent to (30) but are presented in more specialised cross-section or time-series settings. Some of these variants can behave in counterintuitive ways. For example, under the strong disposability assumptions maintained by Färe et al. (1992), the input-based Malmquist productivity index defined in their Eq. (9) implies that productivity increases when inputs increase and outputs fall.

Finally, it is useful to distinguish these Malmquist productivity indexes from the index defined as the ratio of a Malmquist output quantity index to a Malmquist input quantity index (as discussed in Section 5.1). That ratio is commonly known as the Hicks–Moorsteen–Bjurek (HMB) productivity index, defined implicitly or explicitly by Hicks (1961), Moorsteen (1961), Diewert (1992, p. 240), and Bjurek (1996). Except in the restrictive special cases in which the underlying Malmquist quantity indexes are themselves proper, the HMB index is not transitive and therefore, by the criteria established earlier, does not constitute a proper productivity index.

7.3 Numerical Example

This subsection uses a small numerical example to illustrate how different productivity measures behave in practice. Consider an industry in which firms use a single (aggregate) input x to produce two outputs (q_1, q_2). Tables 1 and 2 report data and associated productivity index numbers for four firms operating in this industry; output prices are in the columns labelled p_1 and p_2, output quantities are in the columns labelled q_1 and q_2, and the input quantity is in the column labelled x. All index numbers in Tables 1 and 2 compare each firm with firm A, so the reference observation (row A) takes the value 1 in every index column.

In this example, firm B uses the same input to produce the same output bundle as firm C; any coherent productivity comparison must therefore assign these two firms the same index number. The tables also include firm D, which uses three times the input to produce three times the output of firm A; a coherent comparison must therefore assign firm D the same index number as firm A.

Table 1 reports productivity index numbers constructed as the ratio of a proper output index to a proper input index. Because there is only one input, all input index formulas coincide, so differences in productivity comparisons arise entirely from differences in the output index numbers. The AEW and MEW productivity measures are based on additive and multiplicative output indexes that give the two outputs equal weight. The Lowe and geometric Young (GY) measures are based on additive and multiplicative output indexes that use comparison-invariant weights constructed from average prices and revenue shares, respectively. For simplicity, the primal output index was computed under the assumption that there is no technical or environmental change and the output distance function has the CES form

$$D_O(x, q) = (Ax^{\eta})^{-1} \left(\gamma_1 q_1^{\tau} + \gamma_2 q_2^{\tau}\right)^{1/\tau}, \tag{31}$$

with $A = 1$, $\eta = 0.8$, $\gamma_1 = \gamma_2 = 0.5$ and $\tau = 1.2$. The dual output index was computed using the associated cost function,

$$C(w, q) = A^{-1/\eta}\, w \left(\gamma_1 q_1^{\tau} + \gamma_2 q_2^{\tau}\right)^{1/(\tau\eta)}, \tag{32}$$

where w denotes the price of the aggregate input. The BoD measure is based on a benefit-of-the-doubt output index. The geometric mean of any set of proper indexes is also a proper index, so for illustrative purposes the final column in Table 1 reports the geometric mean of all the other indexes. Observe that all measures in Table 1 behave sensibly: the index number assigned to firm B is equal to the number assigned to firm C, and the number assigned to firm D is equal to the number assigned to firm A. Notice also that in this example, the primal and dual measures coincide; this is because both the output distance function and the associated cost function are separable in the same CES aggregator of outputs.

By contrast, the measures reported in Table 2 do not all behave sensibly. Table 2 reports productivity numbers constructed from binary Fisher (BF), binary Törnqvist

Table 1: Numerical example: proper productivity index numbers

Row	p_1	p_2	q_1	q_2	x	AEW	Lowe	MEW	GY	Primal	Dual	BoD	Mean
A	1	1	1	1	1	1	1	1	1	1	1	1	1
B	4	1	4	1	1	2.50	2.31	2.00	2.14	2.30	2.30	3.00	2.35
C	1	4	4	1	1	2.50	2.31	2.00	2.14	2.30	2.30	3.00	2.35
D	1	3	3	3	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

AEW = additive with equal weights; MEW = multiplicative with equal weights; GY = geometric Young; BoD = benefit-of-the-doubt; Mean = geometric mean.

Table 2: Numerical example: non-proper productivity index numbers

Row	p_1	p_2	q_1	q_2	x	BF	BT	CF	CT	EKS	CCD	MPI	HMB
A	1	1	1	1	1	1	1	1	1	1	1	1	1
B	4	1	4	1	1	2.92	2.72	2.92	2.72	2.54	2.41	2.30	2.30
C	1	4	4	1	1	2.00	2.00	2.92	2.72	2.10	2.07	2.30	2.30
D	1	3	3	3	3	1.00	1.00	1.74	1.61	1.09	1.09	1.55	1.00

BF = binary Fisher; BT = binary Törnqvist; CF = chained Fisher; CT = chained Törnqvist; EKS = Eltetö–Köves–Szulc; CCD = Caves–Christensen–Diewert; MPI = Malmquist productivity index; HMB = Hicks–Moorsteen–Bjurek

(BT), chained Fisher (CF), chained Törnqvist (CT), and the multilateral EKS and CCD output indexes.[10] The table also reports Malmquist productivity index (MPI) and Hicks–Moorsteen–Bjurek (HMB) productivity index numbers. First, observe that the BF, BT, EKS and CCD measures assign different productivity numbers to firms B and C, even though those two firms use the same quantity of input to produce the same output bundle. Second, the chained and multilateral measures fail to preserve proportionality in row D. The failures of the BF and BT indexes stem from the fact that their comparison weights are derived from prices and value shares that vary across comparisons. The failures of the chained and multilateral indexes, in turn, stem from the fact that they are constructed from BF and BT comparisons. The MPI measure also fails to preserve proportionality in row D; in this case the failure reflects both the fact that the MPI is not defined as an output index divided by an input index and the fact that the maintained distance function does not exhibit constant returns to scale. Only the HMB measure behaves sensibly. Observe that it coincides numerically with the primal productivity measure in Table 1. This equivalence is due to the fact that the distance function is separable in the CES aggregator of outputs; it should not be interpreted as a general property of the HMB index, which is not transitive in more general settings.

These results are not numerical curiosities. Even in an extremely simple setting with only one input and two outputs, commonly used productivity measures can generate

[10] For the purpose of computing the chained indexes, the four observations were treated as a time series.

comparisons that depend on comparison-specific weights or chaining paths in ways that violate basic identity and proportionality conditions.

7.4 Why Non-Proper Indexes Persist in Empirical Practice

The preceding sections show that non-proper indexes can generate comparisons that are internally inconsistent even in extremely simple empirical settings. This subsection argues that their continued use is nevertheless common, and that the reasons typically offered in its defence do not withstand close scrutiny.

For every widely used non-proper index there exists a proper alternative that uses the same observable data but combines those data in a different way. Continued reliance on non-proper indexes is therefore best understood not as a response to data limitations, but as a response to inherited conventions and to a set of influential arguments that have been repeatedly asserted in the literature without being subjected to careful measurement-theoretic examination.

The remainder of this subsection considers the most influential objections that have been raised against proper index construction, as well as the most influential motivations offered for using non-proper indexes, explaining in each case why they do not provide a compelling basis for empirical practice.

Objection A (Constant Weights): The easiest way to construct proper quantity and productivity indexes is by using constant (comparison-invariant) weights. Arguments against proper indexes are therefore often framed more narrowly as arguments against constant weights. Such arguments can be traced to influential statements in the index-number literature, including claims that constant-weight comparisons are "unrealistic", "arbitrary" or "absurd". Representative statements include:

> "The only formula which conform perfectly to the circular test are index numbers which have constant weights, . . . But, clearly, constant weighting is not theoretically correct. If we compare 1913 with 1914, we need one set of weights; if we compare 1913 with 1915 we need, theoretically at least, another set of weights . . . Similarly, turning from time to space, an index number for comparing the United States and England requires one set of weights, and an index number for comparing the United States and France requires, theoretically at least, another."

(Fisher, 1922, pp. 274, 275)

and

"The fundamental difficulty is that, in most cases, particularly for geographical comparisons or comparisons between remote points of time, it is absurd to assume constant [weights]."

(Frisch, 1936, p.6)

These statements reflect a concern that weights should adapt to differences in economic structure across time or space. However, two problems arise with this line of reasoning.

First, proper indexes do not, in general, require constant weights: some proper index constructions (including benefit-of-the-doubt approaches) allow weights to vary across observations, albeit in a controlled way. Second, there is no economic principle or theory that requires the analyst to alter weights from one comparison to the next. Allowing weights to vary across comparisons can change the implied ordering of observations in multiple-comparison settings. The numerical example in Section 7.3 illustrates this clearly. When the BF and BT indexes are used, identical output bundles can be assigned different index numbers simply because the weights are comparison-specific and therefore vary across pairs. This is a failure of coherent measurement.

Objection B (Substitution bias). In the index number literature, the term "substitution bias" is commonly used to refer to an alleged systematic distortion in constant-weight measures that arises when relative prices change and economic agents substitute across goods; this claim has been used to motivate the widespread adoption of chained index formulas in official statistics and applied productivity analysis. The concern is that constant-weight indexes do not reflect changes in the relative economic importance of goods. Typical statements are of the form:

"Use of fixed-weighted measures of real GDP and prices for periods other than those close to the base period results in a 'substitution bias' that causes an overstatement of growth for periods after the base year and an understatement of growth for periods before the base year. ... [The] alternative chain-type measure ... provides unbiased estimates of growth."

(Landefeld et al., 1995, pp. 31–32)

From a measurement-theoretic perspective, this substitution-bias argument confuses two distinct objects: *quantity change* and *value change*. Quantity indexes are not intended to track changes in values. They are intended to represent changes in quantities in a way that preserves empirically meaningful relationships. If relative prices change, then value growth will generally diverge from quantity growth; that divergence is not bias, it is simply the difference between measuring *quantities* and measuring *values*. The numerical example in Section 7.3 makes this distinction explicit. The proper measures in Table 1 preserve identity and proportionality exactly: identical bundles receive identical index numbers, and proportional changes in quantities generate proportional index values. By contrast, the chained and multilateral measures in Table 2 violate at least one of these requirements. An index that fails such basic requirements is not correcting bias; it is measuring something other than quantity change.

Objection C ("Strange" Orderings). A further criticism is that proper indexes may yield counterintuitive or "strange" orderings of quantity vectors, particularly in settings with multiple outputs and inputs. For example, Rao (2022, pp. 803, 805) argues that "the Lowe index can lead to strange ordering of output quantity vectors … [and] may lead to counter-intuitive conclusions".[11] In the carefully crafted two-period example considered by Rao, a firm changes its output mix in a setting where productivity losses associated with diseconomies of output substitution exactly offset the productivity gains associated with an outward shift in the production frontier. Consequently, the Lowe index indicates no output (and hence productivity) change over the period. Rao interprets this as a defect, concluding that "a shift in the frontier should imply productivity change driven by technical change. However, the Lowe index shows no productivity change!" (Rao, 2022, p. 806).

From a measurement-theoretic perspective, this criticism faces two difficulties. First, it is not specific to proper index-number formulas. In the two-period setting considered by Rao, analogous orderings can arise with the Laspeyres and Paasche quantity indexes. They will also arise with the binary Fisher index in the special case where prices do not change between the two periods. The example therefore does not identify a property that is peculiar to proper indexes.

Second, the criticism rests on a particular *explanatory* definition of productivity change, under which a frontier shift alone is identified with productivity change. That is, Rao implicitly treats productivity change as being driven solely by technical change (and, potentially, technical efficiency change), as in the Malmquist-type framework of Färe et al. (1994). Under the broader quantity-based definition adopted in this paper, productivity change may depend not only on frontier shifts and technical efficiency change, but also on economies of scale and substitution. In empirical settings where scale or substitution effects are present, productivity change can therefore be small when technical progress and technical efficiency improvements are offset by unfavourable scale or substitution (mix) effects. If an ana-

[11]Recall from Section 4 that the Lowe index is an additive index-number formula that aggregates quantities using a comparison-invariant vector of weights.

lyst wishes to attribute measured productivity change to technical change, technical efficiency change, scale effects or substitution effects, that is a separate problem of *explanation*, to be addressed using economic theory and statistical methods, not by relaxing the axioms required for coherent measurement.

The above objections focus primarily on criticisms directed against proper indexes. A related but distinct set of arguments proceeds in the opposite direction by attempting to provide positive motivations for the continued use of non-proper indexes. The most influential of these are as follows.

Motivation A (Superlative Indexes). Non-proper index-number formulas are often motivated by *exactness* results in the economic approach to index numbers. The basic idea is that certain indexes can be represented as ratios of transformation (or distance) functions belonging to the class of flexible functional forms. Indexes that are exact for a second-order flexible functional form are said to be *superlative*. Because flexible functional forms provide local approximations to an unknown function, these exactness theorems are sometimes interpreted as establishing that non-proper indexes are theoretically grounded and therefore appropriate for applied measurement.

The canonical result is due to Caves et al. (1982b), who show that, under a restrictive set of conditions, the Törnqvist index is exact for a translog transformation function. More recently, Fox (2003) provides an analogous result for Fisher indexes, establishing an exact relationship between the Fisher index and a flexible transformation function of the Diewert (1976) form.

Again, there are two problems with this line of argument. First, exactness results for flexible functional forms are local approximation results. Flexibility is a property defined at a point, and the fact that an index is exact for such a form does not imply that it provides a coherent basis for general multiple comparisons. In particular, ratios of flexible functional forms do not, in general, satisfy the axioms required for proper quantity measurement in multilateral or multitemporal settings. The local nature of these approximation results manifests itself in failures of global consistency conditions such as transitivity. Exactness for a flexible functional form therefore does not resolve the fundamental coherence problems illustrated in the numerical example.

Second, some of the maintained conditions required for exactness are internally inconsistent. For example, Caves et al. (1982b) assume that the output distance (or transformation) function is a translog function that is nondecreasing in outputs. But translog functional forms cannot satisfy global monotonicity restrictions except in degenerate cases: imposing global monotonicity requires all second-order coefficients to be zero, in which case the translog function collapses to a Cobb–Douglas function (a less flexible functional form). When regularity and functional-form restrictions cannot simultaneously be satisfied, appeals to exactness theorems amount to relying on maintained assumptions that cannot jointly hold.

The bottom line is that exactness results provide little guidance for empirical practice in the settings where measurement is most needed. An index does not become coherent because it admits an alternative representation under restrictive assumptions. It is coherent only if it satisfies the axioms required for proper quantity measurement, which ensure that multiple comparisons across time and space are internally consistent and interpretable.

Motivation B (Inertia). Finally, non-proper indexes persist partly because of institutional and historical factors. Statistical agencies and applied researchers face strong incentives to preserve continuity, minimise revisions, and conform to established international standards. They also face incentives to maintain comparability across releases, jurisdictions, and statistical systems. In such environments, index formulas that are familiar, easy to implement, and widely embedded in existing systems can persist even when their limitations are well understood. This helps explain why non-proper indexes remain common in productivity analysis, despite the availability of proper alternatives that use the same underlying data and are often simpler to compute.

Taken together, these arguments suggest that the continued use of non-proper indexes is best explained by convention rather than necessity. Proper indexes do not require additional data, and, in general, they do not require stronger economic assumptions than those implicitly invoked by non-proper approaches. Their main requirement is that empirical comparisons be constructed in a way that is internally consistent and therefore coherent.

7.5 Implementing Proper Productivity Measurement

The preceding subsections show that commonly used non-proper indexes can generate internally inconsistent comparisons, and that many standard arguments used to defend them do not withstand measurement-theoretic scrutiny. The relevant question is therefore not whether coherent productivity measurement is feasible, but how it can be implemented in practice.

For a production economist, a natural route to proper quantity and productivity indexes is via primal or dual representations of production frontiers. In practice, these frontiers are most often estimated using data envelopment analysis (DEA) or stochastic frontier analysis (SFA) methods. These methods are primarily designed to recover production frontiers for explanatory purposes, rather than to deliver aggregation structures suitable for primal or dual index construction. When they are used for index-number construction, certain structural issues arise.

In the case of DEA, the estimated frontier is piecewise linear, with supporting hyperplanes that typically differ across observations. As a result, the frontier is not represented by a single globally defined distance function, but by a collection of linear segments, each defined by its own parameter vector. Because the active supporting hyperplane can vary across observations, the estimated object does not automatically provide a single comparison-invariant aggregation structure suitable for direct primal or dual index construction. Nevertheless, DEA estimators can be used to construct proper quantity and productivity indexes by employing averages of estimated shadow prices or shadow value shares as weights in additive or multiplicative index-number formulas (see, e.g., O'Donnell, 2018, Sec. 6.5).

In the case of SFA, the estimated function is a parametric approximation to an unknown distance, cost, or revenue function. Even if the chosen functional form imposes global structure, it remains an approximation to the true aggregator. This matters because primal and dual indexes are defined conceptually with respect to the true distance, cost, or revenue functions, rather than with respect to approximations to those functions. Nevertheless, SFA estimators can be used to construct proper quantity and productivity indexes by employing the estimated parameters of the approximating function as weights in additive or multiplicative index-number formulas. For this approach to be coherent, the estimated parameters must satisfy the regularity conditions required for economically meaningful aggregation. In practice, this may require imposing nonnegativity constraints during estimation and, when multiplicative forms are used, normalising the parameters so that they sum to one (see, e.g., O'Donnell, 2018, Sec. 8.5).

Additional practical considerations arise in implementation. For example, additive index forms are based on linear aggregator functions and therefore align naturally with DEA-based frontier methods, many of which are formulated as linear programs. Additive indexes also remain well defined when some other index forms are undefined or degenerate (multiplicative index numbers, for example, can become undefined or degenerate when some components take zero or boundary values). Finally, when datasets are small, structured index constructions — including additive indexes — often yield more informative comparisons than benefit-of-the-doubt (BoD) methods (BoD methods may exhibit limited discriminating power in such settings).

Such considerations may influence the preferred *form* of a proper index in a given application, but they do not undermine the measurement principles that define proper index construction.

8 Conclusion

This paper has argued that productivity measurement is fundamentally a problem of quantity measurement. When productivity is defined as the ratio of an output quantity to an input quantity, empirical measurement reduces to the construction of output and input quantity indexes that support coherent comparisons across time and space. Indexes that satisfy a small set of axioms—referred to here as *proper* quantity indexes—are the only measures guaranteed to provide internally consistent representations of productive performance.

The paper clarifies several longstanding sources of confusion in the productivity literature. First, it distinguishes clearly between *measurement* and *explanation*: productivity is an object to be measured prior to, and independently of, any economic model used to explain its variation. Second, it shows that many widely used productivity measures—including chained, multilateral, implicit, and Malmquist-type constructions—do not, in general, satisfy the axioms required for coherent multitemporal and multilateral comparisons. The numerical example demonstrates that the resulting inconsistencies are not merely theoretical possibilities, but can arise even in extremely simple empirical settings. Third, the paper explains why such measures nevertheless persist in applied work, highlighting the roles of inherited conventions, misinterpretations of substitution bias and exactness results, and institutional incentives that favour continuity over coherence.

An important implication is that coherent productivity measurement is already feasible in most empirical contexts. Proper productivity indexes typically require no additional data beyond the quantities, prices, and value shares used to construct non-proper alternatives. What is required instead is a shift in perspective: from viewing productivity measurement as an econometric or modelling exercise to recognising it as a problem of numerical representation of quantity vectors. That shift extends beyond productivity analysis itself. Whenever quantities are aggregated to support comparisons across time or space, the validity of those comparisons depends on the properties of the underlying quantity measures. The framework developed in this paper therefore applies more generally to empirical settings in which quantity aggregation plays a central role.

Once productivity has been measured coherently, analysis of its determinants can proceed on a firm conceptual footing. Economic models, frontier estimation methods, and decomposition frameworks play an essential role in explaining productivity change, but they operate at a logically subsequent stage. Maintaining this distinction allows measurement to remain internally consistent while permitting explanation to focus on technological, environmental, and behavioural influences on productive performance. Detailed treatments of parametric and nonparametric methods for explaining variations in proper productivity index numbers are provided in O'Donnell (2018).

Looking ahead, future work should move beyond treating alternative index formulas as interchangeable empirical tools and instead evaluate productivity measures according to whether they preserve the relationships that productivity comparisons are intended to represent. Advances in productivity analysis are therefore likely to come not only from richer data and more sophisticated econometric methods, but also from closer attention to the principles of coherent measurement.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451. doi:10.3982/ECTA13408.

Aczel, J. and Eichhorn, W. (1974). A note on additive indices. *Journal of Economic Theory*, 8(4):525–529. doi:10.1016/0022-0531(74)90025-8.

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2022). Misallocation, selection, and productivity: A quantitative analysis with panel data from china. *Econometrica*, 90(3):1261–1282. doi:10.3982/ecta16598.

Australian Bureau of Statistics (2021). Australian system of national accounts: Concepts, sources and methods, 2020–21. https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/australian-system-national-accounts-concepts-sources-and-methods/2020-21/chapter-6-price-and-volume-measures/deriving-elemental-volume-estimates. Accessed 28 February 2026.

Australian Bureau of Statistics (2024). Estimates of industry multifactor productivity: Methodology. https://www.abs.gov.au/methodologies/estimates-industry-multifactor-productivity-methodology/2024-25. Accessed 28 February 2026.

Balk, B. (1998). *Industrial Price, Quantity, and Productivity Indices: The Micro-Economic Theory and an Application.* Kluwer Academic Publishers, Boston. doi:10.1007/978-1-4757-5454-4.

Ball, E., Wang, S., Nehring, R., and Mosheim, R. (2016). Productivity and economic growth in U.S. agriculture: A new look. *Applied Economic Perspectives and Policy*, 38(1):30–49. doi:10.1093/aepp/ppv031.

Barro, R. J. (1999). Notes on growth accounting. *Journal of Economic Growth*, 4(2):119–137. doi:10.1023/a:1009828704275.

Basu, S., Fernald, J., and Shapiro, M. (2001). Productivity growth in the 1990s: technology, utilization, or adjustment? *Carnegie-Rochester Conference Series on Public Policy*, 55:117–165. doi:10.1016/s0167-2231(01)00054-9.

Bjurek, H. (1996). The Malmquist total factor productivity index. *The Scandinavian Journal of Economics*, 98(2):303–313. doi:10.2307/3440861.

Boussemart, J.-P., Briec, W., Kerstens, K., and Poutineau, J.-C. (2003). Luenberger and Malmquist productivity indices: Theoretical comparisons and empirical illustration. *Bulletin of Economic Research*, 55(4):391–405. doi:10.1111/1467-8586.00183.

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982a). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 92(365):73–96. doi:10.2307/1913388.

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982b). Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal*, 50(6):1393–1414. doi:10.2307/2232257.

Chambers, R., Färe, R., and Grosskopf, S. (1996). Productivity growth in APEC countries. *Pacific Economic Review*, 1(3):181–190. doi:10.1111/j.1468-0106.1996.tb00184.x.

Chambers, R. and Pope, R. (1996). Aggregate productivity measures. *American Journal of Agricultural Economics*, 78(5):1360–1365. doi:10.2307/1243522.

Cherchye, L., Moesen, W., Rogge, N., and van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research*, 82(1):111–145. doi:10.1007/s11205-006-9029-7.

Christensen, L. (1975). Concepts and measurement of agricultural productivity. *American Journal of Agricultural Economics*, 57(5):910–915. doi:10.2307/1239102.

Chung, Y., Färe, R., and Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51(3):229–240. doi:10.1006/jema.1997.0146.

Coelli, T. and Rao, D. (2005). Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980-2000. *Agricultural Economics*, 32(s1):115–134. doi:10.1111/j.0169-5150.2004.00018.x.

Diewert, W. (1992). Fisher ideal output, input, and productivity indexes revisited. *Journal of Productivity Analysis*, 3(3):211–248. doi:10.1007/bf00158354.

Diewert, W. and Fox, K. (2018). Decomposing value-added growth into explanatory factors. In Grifell-Tatjé, E., Lovell, C., and Sickles, R., editors, *The Oxford Handbook of Productivity Analysis*, chapter 19, pages 625–662. Oxford University Press, New York. doi:10.1093/oxfordhb/9780190226718.013.19.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4(2):115–145. doi:10.1016/0304-4076(76)90009-9.

Drechsler, L. (1973). Weighting of index numbers in multilateral international comparisons. *Review of Income and Wealth*, 19(1):17–34. doi:10.1111/j.1475-4991.1973.tb00871.x.

Eichhorn, W. (1976). Fisher's tests revisited. *Econometrica*, 44(2):247–256. doi:10.2307/1912721.

Epure, M., Kerstens, K., and Prior, D. (2011). Bank productivity and performance groups: A decomposition approach based upon the Luenberger productivity indicator. *European Journal of Operational Research*, 211(3):630–641. doi:10.1016/j.ejor.2011.01.041.

Eurostat (2020). Building the system of national accounts: Volume measures. https://ec.europa.eu/eurostat/statistics-explained/index.php/Building_the_System_of_National_Accounts_-_volume_measures. Accessed: 28 February 2026.

Färe, R. and Grosskopf, S. (1996). *Intertemporal Production Frontiers: With Dynamic DEA*. Springer Dordrecht. doi:https://doi.org/10.1007/978-94-009-1816-0.

Färe, R., Grosskopf, S., Lindgren, B., and Roos, P. (1992). Productivity changes in Swedish pharmacies 1980-1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3:85–101. doi:10.1007/bf00158770.

Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1):66–83. URL https://www.jstor.org/stable/2117971.

Färe, R. and Zelenyuk, V. (2021). On aggregation of multi-factor productivity indexes. *Journal of Productivity Analysis*, 55(2):107–133. doi:10.1007/s11123-021-00598-w.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the Penn World Table. *The American Economic Review*, 105(10):3150–3182. doi:10.1257/aer.20130954.

Feng, G. and Serletis, A. (2010). A primal Divisia technical change index based on the output distance function. *Journal of Econometrics*, 159(2):320–330. doi:10.1016/j.jeconom.2010.09.006.

Feroze, S. and Chauhan, A. (2010). Performance of dairy self help groups (SHGs) in India: Principal component analysis (PCA) approach. *Indian Journal of Agricultural Economics*, 65(2):308–320. doi:10.22004/ag.econ.204684.

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

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425. doi:10.1257/aer.98.1.394.

Fox, K. (2003). An economic justification for the EKS multilateral index. *Review of Income and Wealth*, 49(3):407–413. doi:10.1111/1475-4991.00095.

Frisch, R. (1936). Annual survey of general economic theory: The problem of index numbers. *Econometrica*, 4(1):1–38. doi:10.2307/1907119.

Garcia-Marin, A. and Voigtländer, N. (2019). Exporting and plant-level efficiency gains: It's in the measure. *Journal of Political Economy*, 127(4):1777–1825. doi:10.1086/701607.

Grifell-Tatjé, E. and Lovell, C. (1995). A note on the Malmquist productivity index. *Economics Letters*, 47:169–175. doi:10.1016/0165-1765(94)00497-p.

Harris, R. and Moffat, J. (2015). Plant-level determinants of total factor productivity in Great Britain, 1997–2008. *Journal of Productivity Analysis*, 44(1):1–20. doi:10.1007/s11123-015-0442-2.

Hicks, J. (1961). *Measurement of Capital in Relation to the Measurement of Other Economic Aggregates*. Macmillan, London. doi:10.1007/978-1-349-08452-4_2.

Hill, P. (2008). Lowe indices. Paper presented at the 2008 World Congress on National Accounts and Economic Performance Measures for Nations, Washington, D.C.

Hranaiova, J. and Stefanou, S. (2002). Scale, productivity growth and risk response under uncertainty. *Economia Agraria y Recursos Naturales*, 2(2):73–91. doi:10.7201/earn.2002.02.04.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in china and india. *Quarterly Journal of Economics*, 124(4):1403–1448. doi:10.1162/qjec.2009.124.4.1403.

Jorgenson, D. W. and Griliches, Z. (1967). The explanation of productivity change. *Review of Economic Studies*, 34(3):249–283. doi:10.2307/2296675.

Klette, T. J. (1999). Market power, scale economies and productivity: Estimates from a panel of establishment data. *Journal of Industrial Economics*, 47(4):451–476. doi:10.1111/1467-6451.00108.

Kuosmanen, T. and Sipiläinen, T. (2009). Exact decomposition of the Fisher ideal total factor productivity index. *Journal of Productivity Analysis*, 31(3):137–150. doi:10.1007/s11123-008-0129-z.

Landefeld, J., Parker, R., and Triplett, J. (1995). Preview of the comprehensive revision of the national income and product accounts: BEA's new featured measures of output and prices. *Survey of Current Business*, 75(July):31–38.

McCormack, M., Thorne, F., and Hanrahan, K. (2020). Measuring total factor productivity on Irish dairy farms: a Fisher index approach using farm-level data. *Irish Journal of Agricultural and Food Research*, 59(1):123–159. doi:10.15212/ijafr-2020-0107.

Moorsteen, R. (1961). On measuring productive potential and relative efficiency. *The Quarterly Journal of Economics*, 75(3):151–167. doi:10.2307/1885133.

Nagar, A. and Basu, S. (2002). Weighting socioeconomic indicators of human development: A latent variable approach. In Ullah, A., Wan, A., and Chaturvedi, A., editors, *Handbook of Applied Econometrics and Statistical Inference*, volume 165, chapter 29, pages 609–641. Marcel Dekker, New York.

O'Donnell, C. (2012). Nonparametric estimates of the components of productivity and profitability change in U.S. agriculture. *American Journal of Agricultural Economics*, 94(4):873–890. doi:10.1093/ajae/aas023.

O'Donnell, C. J. (2016). Using information about technologies, markets and firm behaviour to decompose a proper productivity index. *Journal of Econometrics*, 190(2):328–340. doi:10.1016/j.jeconom.2015.06.009.

O'Donnell, C. J. (2018). *Productivity and Efficiency Analysis: An Economic Approach to Measuring and Explaining Managerial Performance.* Springer Nature, Singapore. doi:10.1007/978-981-13-2984-5.

OECD (2001). *OECD Manual: Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth.* Organisation for Economic Cooperation and Development, Paris. doi:10.1787/9789264194519-en.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297. doi:10.2307/2171831.

O'Mahony, M. and Timmer, M. (2009). Output, input and productivity measures at the industry level: The EU KLEMS database. *Economic Journal*, 119(538):F374–F403. doi:10.1111/j.1468-0297.2009.02280.x.

Pan, M. and Walden, J. (2015). Measuring productivity in a shared stock fishery: A case study of the Hawaii longline fishery. *Marine Policy*, 62:302–308. doi:10.1016/j.marpol.2015.07.018.

Petrin, A. and Levinsohn, J. (2012). Measuring aggregate productivity growth using plant-level data. *The Rand Journal of Economics*, 43(4):705–725. doi:10.1111/1756-2171.12005.

Rao, D. (2022). Index numbers and productivity measurement. In Ray, S., Chambers, R., and Kumbhakar, S., editors, *Handbook of Production Economics*, chapter 19. Springer Nature, Singapore. doi:10.1007/978-981-10-3455-8_8.

Ray, S. and Mukherjee, K. (1996). Decomposition of the Fisher ideal index of productivity: A non-parametric dual analysis of US airlines data. *The Economic Journal*, 106(439):1659–1678. doi:10.2307/2235206.

Rogge, N. (2018). On aggregating benefit of the doubt composite indicators. *European Journal of Operational Research*, 264:364–369. doi:10.1016/j.ejor.2017.06.035.

Salim, R. and Islam, N. (2010). Exploring the impact of R&D and climate change on agricultural productivity growth: the case of Western Australia. *Australian Journal of Agricultural and Resource Economics*, 54(4):561–582. doi:10.1111/j.1467-8489.2010.00514.x.

Schreyer, P. (2010). Measuring multi-factor productivity when rates of return are exogenous. In Diewert, W. E., Balk, B. M., Fixler, D., Fox, K. J., and Nakamura, A. O., editors, *Price and Productivity Measurement: Volume 6 – Index Number Theory*, chapter 2, pages 13–40. Trafford Press.

Sickles, R. and Zelenyuk, V. (2019). *Measurement of Productivity and Efficiency: Theory and Practice.* Cambridge University Press, Cambridge, UK. doi:10.1017/9781139565981.

Skevas, I. (2023). A novel modeling framework for quantifying spatial spillovers on total factor productivity growth and its components. *American Journal of Agricultural Economics*, 105:1221–1247. doi:10.1111/ajae.12360.

Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics*, 39(3):312–320. doi:10.2307/1926047.

Star, S. (1974). Accounting for the growth of output. *The American Economic Review*, 64(1):123–135.

Swamy, S. (1965). Consistency of Fisher's tests. *Econometrica*, 33(3):619–623. doi:10.2307/1911757.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2):326–365. doi:10.1257/jel.49.2.326.

USDA (2023). Productivity growth in U.S. agriculture. https://www.ers.usda.gov/data-products/agricultural-productivity-in-the-united-states/productivity-growth-in-us-agriculture. Accessed 27 February 2026.

van Puyenbroeck, T. and Rogge, N. (2017). Geometric mean quantity index numbers with benefit-of-the-doubt weights. *European Journal of Operational Research*, 256:1004–1014. doi:10.1016/j.ejor.2016.07.038.

Zelenyuk, V. and Zhao, S. (2025). Inference for DEA estimators of Malmquist productivity indices: an overview, further improvements, and a guide for practitioners. *Macroeconomic Dynamics*, 29(e81):1–22. doi:10.1017/s1365100525000094.

Appendix: Divisia Decompositions

This appendix derives the Divisia-type local productivity-growth expressions (26) and (28) in the text. We begin with the derivation of (26).

By the chain rule,

$$\frac{d}{dt}\ln D_O^t\big(x(t),q(t),z(t)\big) = \sum_{n=1}^{N}\frac{\partial \ln D_O^t(x,q,z)}{\partial q_n}\frac{dq_n(t)}{dt} + \sum_{m=1}^{M}\frac{\partial \ln D_O^t(x,q,z)}{\partial x_m}\frac{dx_m(t)}{dt} + \sum_{j=1}^{J}\frac{\partial \ln D_O^t(x,q,z)}{\partial z_j}\frac{dz_j(t)}{dt} + \frac{\partial \ln D_O^t(x,q,z)}{\partial t},$$

where all partial derivatives on the right-hand side are evaluated at $(x,q,z,t)=(x(t),q(t),z(t),t)$. Rewriting each term on the right-hand side using logarithmic derivatives yields

$$\frac{d}{dt}\ln D_O^t\big(x(t),q(t),z(t)\big) = \sum_{n=1}^{N}\frac{\partial \ln D_O^t}{\partial \ln q_n}\dot{q}_n(t) + \sum_{m=1}^{M}\frac{\partial \ln D_O^t}{\partial \ln x_m}\dot{x}_m(t) + \sum_{j=1}^{J}\frac{\partial \ln D_O^t}{\partial \ln z_j}\dot{z}_j(t) + \frac{\partial \ln D_O^t}{\partial t}.$$

After some simple rearrangement, this yields (26), where $\phi_n(t) \equiv \partial \ln D_O^t/\partial \ln q_n$, $\gamma_m(t) \equiv -\partial \ln D_O^t/\partial \ln x_m$, $TP_O(t) \equiv -\partial \ln D_O^t/\partial t$, $EC_O(t) \equiv -\sum_{j=1}^{J}(\partial \ln D_O^t/\partial \ln z_j)\dot{z}_j(t)$, and $OTEC(t) \equiv \frac{d}{dt}\ln D_O^t(x(t),q(t),z(t))$.

We now derive equation (28). This decomposition involves four main steps.

1. By the chain rule,

$$\frac{d}{dt}\ln C^t\big(w(t),q(t),z(t)\big) = \sum_{m=1}^{M}\frac{\partial \ln C^t(w,q,z)}{\partial w_m}\frac{dw_m(t)}{dt} + \sum_{n=1}^{N}\frac{\partial \ln C^t(w,q,z)}{\partial q_n}\frac{dq_n(t)}{dt} + \sum_{j=1}^{J}\frac{\partial \ln C^t(w,q,z)}{\partial z_j}\frac{dz_j(t)}{dt} + \frac{\partial \ln C^t(w,q,z)}{\partial t},$$

   where all partial derivatives on the right-hand side are evaluated at $(w,q,z,t) = (w(t),q(t),z(t),t)$. Rewriting each term using logarithmic derivatives yields

$$\frac{d}{dt}\ln C^t\big(w(t),q(t),z(t)\big) = \sum_{m=1}^{M}\frac{\partial \ln C^t}{\partial \ln w_m}\dot{w}_m(t) + \sum_{n=1}^{N}\frac{\partial \ln C^t}{\partial \ln q_n}\dot{q}_n(t) + \sum_{j=1}^{J}\frac{\partial \ln C^t}{\partial \ln z_j}\dot{z}_j(t) + \frac{\partial \ln C^t}{\partial t}. \quad (33)$$

2. Define cost efficiency as $CE(t) \equiv C^t(w(t),q(t),z(t))/C(t)$ where $C(t) \equiv w(t)'x(t)$

is observed cost. Taking logarithms and differentiating yields

$$\frac{d}{dt}\ln CE(t) = \frac{d}{dt}\ln C^t\big(w(t), q(t), z(t)\big) - \frac{d}{dt}\ln C(t).$$

Equivalently,

$$\frac{d}{dt}\ln C(t) = \frac{d}{dt}\ln C^t\big(w(t), q(t), z(t)\big) - \frac{d}{dt}\ln CE(t). \tag{34}$$

Substituting (33) into (34) yields

$$\begin{aligned}\frac{d}{dt}\ln C(t) = &\sum_{m=1}^{M}\frac{\partial \ln C^t}{\partial \ln w_m}\,\dot{w}_m(t) + \sum_{n=1}^{N}\frac{\partial \ln C^t}{\partial \ln q_n}\,\dot{q}_n(t)\\ &+\sum_{j=1}^{J}\frac{\partial \ln C^t}{\partial \ln z_j}\,\dot{z}_j(t) + \frac{\partial \ln C^t}{\partial t} - \frac{d}{dt}\ln CE(t).\end{aligned} \tag{35}$$

3. Differentiating $C(t) = \sum_{m=1}^{M} w_m(t)x_m(t)$ yields

$$\frac{dC(t)}{dt} = \sum_{m=1}^{M} x_m(t)\frac{dw_m(t)}{dt} + \sum_{m=1}^{M} w_m(t)\frac{dx_m(t)}{dt}.$$

Dividing by $C(t)$ yields

$$\frac{d}{dt}\ln C(t) = \sum_{m=1}^{M}\frac{x_m(t)}{C(t)}\frac{dw_m(t)}{dt} + \sum_{m=1}^{M}\frac{w_m(t)}{C(t)}\frac{dx_m(t)}{dt}.$$

Re-expressing each term using logarithmic derivatives gives

$$\frac{d}{dt}\ln C(t) = \sum_{m=1}^{M} s_m(t)\,\dot{w}_m(t) + \sum_{m=1}^{M} s_m(t)\,\dot{x}_m(t), \tag{36}$$

where $s_m(t) \equiv w_m(t)x_m(t)/C(t)$.

4. Equating (35) and (36) and rearranging yields (28), where $\beta_n(t) \equiv \partial \ln C^t/\partial \ln q_n$, $\alpha_m(t) \equiv \partial \ln C^t/\partial \ln w_m$, $TP_C(t) \equiv -\partial \ln C^t/\partial t$, $EC_C(t) \equiv -\sum_{j=1}^{J}(\partial \ln C^t/\partial \ln z_j)\dot{z}_j(t)$, and $CEC(t) \equiv d\ln CE(t)/dt$.