Microfoundations in measurement and theory[1]

Junyi Zhu

Deutsche Bundesbank

2025.02.28

Abstract

This paper clarifies that, to address the aggregation problem, macroeconomic models must return to the empirical aggregates of national accounts (NA), thereby reconciling theoretical constructs with observed macro outcomes. However, some fundamental micro–macro gaps emerge between the bottom-up, microfounded macro and macro measurement. The primary gap stems from the divergence between closed aggregation—rooted in micro reality and theory—and the open aggregation approach employed by NA, which implicitly imposes macro restrictions that lack full microfoundations (Schlicht, 1985). In practice, two barriers—information friction and policy-oriented objectives—prevent macro measurement from adopting a closed aggregation framework. Using the aggregate production function and the contrast between Real GDP and Physical Value Added as illustrative examples, we demonstrate these barriers and rationalize NA's reliance on open aggregation. A microfounded explanation is achievable only when the corresponding micro restrictions are satisfied, ensuring consistency between bottom-up aggregation and macro measurement; otherwise, NA aggregates diverge from those derived from micro data. Recognizing these gaps, we argue that many current efforts to redistribute the aggregate data exhibit a paradoxical regression to representative agent thinking. Tentative solutions include the development of satellite accounts and/or the imposition of top-down micro–macro restrictions—drawing on index number theory and practice—on the bottom-up aggregation process.

Keywords: micro-macro gap; aggregation problem; index number

JEL-Classification: C43, E1, E01, B4

[1] E-Mail: junyi.zhu@bundesbank.de. The paper represents the author's personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

1 Introduction

The growing accessibility of granular microdata together with mounting concerns about inequality has spurred interest in redistributing National Accounts (NA), offering a timely opportunity to re-examine the long-neglected aggregation problem. At its core, this endeavor for breaking down aggregate data stems from the need, shared by the aggregation problem, to ensure consistency between the macro relationships published by national statistical offices and their bottom-up counterparts derived from microdata. This challenge is often termed *closing the micro-macro (data) gap*. Yet this task exposes a deeper dual challenge in both aggregation in microfoundation theory and disaggregation in measurement.

A paradox arises. Where is the challenge: if NA aggregates are inherently bottom-up constructions, why does redistributing them—by reversing the aggregation process—fail to align perfectly with microdata? By this logic, shouldn't the aggregation problem diminish as microdata accessibility and micro modeling improve? After all, bottom-up aggregation (as practiced in NA) must yield macro relationships consistent with macro realities measured in NA from the same micro reality? In this view, aggregation in theory must coincide with aggregation in measurement. Yet as Klein (1946a) noted in the seminal microfoundations debate:

The problem of bridging the gap between the traditional theories based on individual behavior and the theories based on community or class behavior is, to a large extent, a problem of proper measurement.

In this paper, we extend Klein's insight by formalizing a critical distinction: *closed aggregation* (theory-driven, micro-consistent) versus *open aggregation* (measurement-driven, macro-constrained), following Schlicht (1995).[2] These divergent approaches produce irreconcilable macro relationships even under identical micro realities. Closed aggregation enforces strict adherence to microfoundational principles, while open aggregation prioritizes macro-level coherence, often at the expense of micro fidelity. This dichotomy explains why redistributing NA—a process theoretically akin to reversing aggregation—fails to resolve the micro-macro gap: measurement constraints and theoretical assumptions are not merely inverses but *competing frameworks*.

[2] Throughout this paper, we use the terms (macro) measurement and open aggregation interchangeably, while micro theory-based macro (or microfounded macro) and closed aggregation are treated as synonymous.

To illustrate, consider an economy composed of two individuals, A and B. Their micro-level variables, denoted by $x = \{x_1, x_2\ x_3, x_4\}$, are governed by the micro restrictions, which are partitioned into two sets

$$E_{f^1} \coloneqq \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0, x_2 = 0\} \quad (1)$$
$$E_{f^2} \coloneqq \{(x_3, x_4) \in \mathbb{R}_+^2 \mid x_3 = |x_1|, x_4 = |x_2|)\} \quad (2)$$

Equation (1) describes their endowment: x_1 and x_2 represent incomes for persons A and B, respectively. Their choice is characterized by equation (2): x_3 and x_4 denote consumptions for A and B.[3] The aggregation function $\alpha(x)$ maps micro-level variables to the macro-level variables $X = \{X^1, X^2\}$, where X^1 represents the aggregate income and X^2 represents the aggregate consumption

$$X^1 = x_1 + x_2 \quad (3)$$
$$X^2 = x_3 + x_4 \quad (4)$$

Under closed aggregation, micro restrictions are applied *before* aggregation (*restrict then aggregate*). The resulting macro equilibrium is

$$E_F \coloneqq \alpha(E_{f^1} \cap E_{f^2}) \quad (5)$$

This conforms to our conventional bottom-up intuition and the neoclassical microfoundation theory, yielding a unique macro relationship:

$$E_F = \{X^1, X^2\} = \{0,0\} \quad (6)$$

Under open aggregation, micro-level variables are aggregated *within each partition independently*, before identifying the macro-level relationships (*aggregate then restrict*):

$$E_{F^1}^o = \{X^1 = 0\} \quad (7)$$
$$E_{F^2}^o = \{X^2 \in \mathbb{R}_+\} \quad (8)$$

Then the macro restrictions are then combined:

$$E_F^o \coloneqq E_{F^1}^o \cap E_{F^2}^o \quad (9)$$

This results in a broader macro equilibrium:

$$E_F^o = \{X^1, X^2\} = \{0, \mathbb{R}_+\} \quad (10)$$

We argue the open aggregation characterizes the macro measurement. Many aggregates in the National Accounts (NA) are the synthetic ones, which are constructed by using

[3] Their incomes can be negative, implying they can borrow from or be levied by the third party.

the macro relationships, such as (10), specific to the open aggregation.[4] Their coherence emerges from system-wide constraints or policy-driven definitions rather than strict micro-to-macro consistency like (6), while those natural aggregates, like consumption and income in our example, should be consistent between theory and measurement in principle.

To challenge the reductionist assumption that NA aggregates directly mirror micro-level realities, consider the measurement of real value added. While nominal value added—calculated as total output minus intermediate inputs—holds residually at both aggregate and subaggregate levels (e.g., industries or firms), NA employs *double deflation* to preserve this identity in the real terms, but exclusively at the macro level. Here, aggregate real value added is derived as the residual between deflated total output and deflated intermediate inputs, with separate deflators applied to each component. Yet this macro-level coherence collapses at subaggregate levels: applying deflators identical to the macro ones frequently fails to reconcile real value added with its output-input residual.[5]

This discrepancy underscores a foundational tension in economic measurement: NA deflators are engineered to enforce macro-level consistency, prioritizing aggregate relationships over micro-level fidelity. The resulting index numbers reflect macro price/quantity dynamics—*not* their micro counterparts—revealing a pragmatic, top-down logic beneath NA's ostensibly "bottom-up" nominal aggregates. In essence, NA's real-terms construction sacrifices micro-theoretical purity to uphold macro coherence, exposing the limits of reductionist approaches to aggregation.

This paper addresses the persistent micro–macro gap by identifying and reconciling two categorically distinct domains that contribute to this divergence. Crucially, we shift the focus from the errors in micromeasurement and classification to the discrepancy inherent in the aggregation methodology. Specifically, NA must impose independent macro restrictions (Schlicht, 1985) to capture aggregate outcomes, as closed aggregation is often infeasible. In doing so, we reveal two sides of the same coin:

- The difficulty of forming microfounded explanations when synthetic aggregates (e.g., real GDP, productivity indices) rely on unmet micro restrictions;
- The compromised alignment between bottom-up aggregated macro data and NA-measured outcomes—the micro-macro data gap.

[4] See Hoover (1995) for the distinction between natural and synthetic aggregates, which parallels Kant's analytic-synthetic dichotomy.

[5] The exercise here is to propagate all the three macro deflators to any subaggregate unit after constructing the value-added deflator through double deflation at the macro level.

By reframing the aggregation problem as a translational challenge between economists and NA statisticians, this paper underscores the necessity of "taking measurement seriously". Rather than dismissing NA constructs as ad hoc, we argue that their open aggregation logic reflects unavoidable compromises in reconciling micro heterogeneity with macro coherence. Conversely, economists' insistence on closed aggregation—whether implicit in representative agent models or explicit in heterogeneous agent frameworks—obscures critical measurement constraints, perpetuating mutual misunderstanding. To bridge this divide, we propose macroeconomic models that balance rigorous microfounded principles with empirical realities of national accounts. Tentative solutions include developing satellite accounts and imposing top-down micro–macro restrictions.

Given these challenges, shouldn't we begin a more explicit debate on the ignored role of measurement in aggregation problem as well as macroeconomic modelling? Incorporating insights from theory of measurement — particularly how national accounts are constructed versus how economic theory treats aggregation — could help bridge the gap between micro and macro realities, leading to models that are both theoretically sound and empirically coherent.

The remainder of the paper proceeds as follows. Section 2 establishes the essential conceptual framework for understanding aggregation problems. Section 3 distinguishes between *closed* (micro-consistent) and *open* (macro-constrained) approaches, aided by an illustration from production function aggregation. Section 4 presents a case study contrasting Real Value Added (a NA staple) with Physical Value Added (a micro-theoretical ideal), revealing how measurement protocols shape—and sometimes distort—economic narratives. Section 5 identifies the fundamental barriers to aggregation, inherent in the impossibility to access a complete micro information. Section 6 critiques the conflation of aggregation in measurement (e.g., the practice of redistributing NA) with representative agent constructs, arguing that this conflation perpetuates theoretical overreach. Finally, Section 7 evaluates tentative solutions, including satellite accounts and top-down restrictions, while Section 8 concludes with implications for policy and research.

2 Essential understanding of aggregation problem

The aggregation is the sine qua non of microfoundations. It is therefore imperative to delineate the aggregation problem in macroeconomics and emphasize the pivotal role played by measurement.

2.1 A definition of aggregation problem

We present such a general exposition which also appears in Schlicht (1985), Janssen (2005) and Van Daal and Merkies (1988). Consider an economy composed of n individuals. There exist the individual relationships between m explanatory variables x_{ij}, $j = 1, \dots, m$, and one dependent variable y_i

$$y_i = f_i(x_{i1}, \dots, x_{im}),\ i = 1, \dots, n \tag{11}$$

The macro world is built by the aggregating functions to link micro and macro

$$Y := G(y_1, \dots, y_n),\ i = 1, \dots, n \text{ and } X_j := g_j(x_{1j}, \dots, x_{nj}), j = 1, \dots, m \tag{12}$$

such that the macro reality Y and X_j are thereby defined. Macroeconomists are intrinsically interested in the macro relationship

$$Y := F(X_1, \dots, X_m) \tag{13}$$

To microfound such a function (13), there are basically two approaches. One is to firstly impose the restrictions (11) connecting the micro explanatory and dependent variables and then aggregate the micro dependent variable (*restrict then aggregate*)

$$\begin{aligned} Y &:= G\big(f_1(x_{11}, \dots, x_{1m}), \dots, f_n(x_{n1}, \dots, x_{nm})\big) \\ &= G'(x_{11}, \dots, x_{1m}, \dots, x_{n1}, \dots, x_{nm}) \end{aligned} \tag{14}$$

Alternatively, people aggregate the micro explanatory and dependent variables before the formation of the macro relationship (*aggregate then restrict*)

$$\begin{aligned} Y &:= F\big(g_1(x_{11}, \dots, x_{n1}), \dots, g_m(x_{1m}, \dots, x_{nm})\big) \\ &= F'(x_{11}, \dots, x_{n1}, \dots, x_{1m}, \dots, x_{nm}) \end{aligned} \tag{15}$$

According to Schlicht (1985), (14) and (15) are respectively classified to be *closed* and *open aggregation*. Note $G(\,)$ in (14) aggregates the *a posteriori* y_i, which is the equilibrium micro outcomes after *jointly* imposing the restrictions from (11). The aggregation problem explores the conditions — particularly those governing the choice of aggregating functions (12) and macro relationship (13) — under which the two approaches of aggregation (14) and (15) are consistent, i.e. $G'(\,) = F'(\,)$.

Take Klein (1946a) for example, m is simply one and we study how to explain the consumption by income. Restrictions (11) are simply the individual consumption functions, which, under the neoclassical framework, are generated by the postulates of rational choice and general equilibrium. Macro relationship (13) is the aggregate demand function mapping between aggregate income and aggregate expenditure.

It is worth emphasizing that we focus on the models enhancing our understanding of the real world. Consequently restrictions imposed in (11) and (13) should reflect the empirical regularities at both micro and macro levels. Such models are indispensable for practical applications like policy analysis and forecasting—it is difficult to envision a macro model failing to align with aggregate statistical relationships yet remaining useful for policymaking. Given economics' pragmatic orientation, ontological individualism (the philosophical bedrock of microfounded macroeconomics) ensures economists rarely adopt micro models divorced from observed micro realities. This underscores why macro restrictions like (13) are often imposed a priori: they emerge from measurement requirements rather than logical aggregation of micro-level restrictions $f_i(\,)$. In this light, $F(\,)$ in (13) is not a mere outcome of aggregating microdata but a construct shaped by empirical necessity. The aggregation problem thus becomes a consistency issue—one that exposes the tension between measurement-driven macro constraints and theory-driven microfoundations.

2.2 Macro restrictions $F(\,)$

A fundamental concern for policymakers and the public lies in understanding and interpreting the statistical relationships that shape economic decision-making. For instance, the mandates of central banks (CBs)—such as controlling the price stability and economic growth—are often evaluated and communicated through macroeconomic indicators, particularly the relationship between aggregate demand and general price level dynamics. Central banks rely on these statistical relationships to inform monetary policy decisions, adjusting interest rates, money supply, or other instruments to maintain economic stability. However, the reliability of these relationships can be challenged by structural changes, policy expectations, and evolving market conditions, making it crucial to continuously assess their empirical validity. Consequently, both policymakers and the public must critically engage with these macroeconomic aggregates to ensure that policy objectives align with real economic conditions.

The role of measurement is pivotal in determining the functional form of $F(\,)$ in macroeconomic modeling. Many macroeconomic relationships, especially in microfounded models, rely on synthetic aggregates that are themselves constructed using non-microfounded statistical restrictions. A prime example is real GDP, which serves as a cornerstone for economic analysis but is derived from national accounting conventions rather than micro-level optimization. Despite advancements in microfounded macroeconomics, traditional time-series regressions, such as Cobb-Douglas functional forms that emerged simply from aggregate-level data fitting, remain widely used in

central banks (CBs) for policy modeling and forecasting.[8] This continued reliance reflects both empirical stability and the enduring predictive power of these statistical formulations.

Moreover, several macroeconomic theories are historically rooted in statistical regularities rather than clear microfoundations. For instance, theories of international trade, monetary dynamics, business cycles, and Keynesian unemployment (as noted by Schlicht, 1985) emerged primarily from observed macro-level patterns rather than first principles of individual behavior. This highlights an ongoing tension in macroeconomics: while modern approaches emphasize deriving macro laws from microeconomic foundations, policymakers and practitioners often still rely on aggregate statistical relationships that lack explicit micro-level justification.

The history of national accounts can be likened to the development and continued use of the conventional thermometer in meteorology. Long before the ideal gas law was reduced to kinetic theory, thermometers were already in widespread use for measuring temperature. Even with later progress in particle physics, the traditional thermometer remains a practical and indispensable tool for national weather services today. Similarly, national accounts and aggregate macroeconomic indicators—such as GDP, inflation, and employment statistics—were developed before the emergence of microfoundations in economic theory. Yet, despite no fundamental change in their construction methodology, they continue to play a central role in policymaking and economic analysis, as they provide reliable, empirically tested measures for understanding the macroeconomy. At the same time, their micro-theoretic underpinnings have continued to evolve.[9] This analogy underscores the persistent value of empirical methods and the role of historical conventions in maintaining the critical status of macroeconomic measurement in modern policymaking.

[8] This case bears two implications: 1) many macro relationships (targeted by microfounded studies) are identified through the data fitting in the macro data; 2) they are not adequately (even less convincingly) explained under closed aggregation – eg Cobb-Douglas function is more plausibly explained through accounting identities (Balk, 2024; Shaikh, 1974). Fisher (1971)'s simulation reveals the causal explanation for the Cobb-Douglas form in the aggregate should come from, instead of the microfoundation, the constant wage share associated with the underlying accounting relation distributing value added into labor and capital input cost in the nominal term.

[9] The theorization of national accounts and the liquid-in-glass thermometer share an inductive foundation, both emerging from empirical regularities rather than fundamental theory. Thermometers were developed based on the observable expansion of liquids with temperature (Boyle's Law) long before kinetic theory provided a deeper explanation. Likewise, GDP, inflation, and employment statistics were constructed from statistical consistency, predating microfounded macroeconomics. The sustained reliance on both highlights the successful application of inductive reasoning in measurement.

Understanding how macroeconomic restrictions are defined a priori based on macro measurement, from which the aggregation problem ensues, helps clarify key debates within central banks, especially in shaping policy recommendations. Macro policy institutes like CBs often exhibit segmentation in their approach to data and models, reflecting contrasting emphases between macro- and micro-level analysis as well as differences in policy objectives. For instance, financial stability assessments typically rely on macroeconomic aggregates and systemic risk models, while bank supervision focuses on microprudential indicators, firm-level balance sheets, and institution-specific risks. This division can lead to challenges in ensuring coherence between macro and micro perspectives, particularly when policy decisions require an integrated view of financial systems.

The Bernanke Review (Bernanke, 2024), commissioned by the Bank of England in 2022 and led by former Federal Reserve Chair Ben Bernanke, critically assessed the Bank's forecasting and policy analysis frameworks. A key takeaway from the review was the need to improve the consistency between data and models, particularly in monetary policy and financial stability analysis. One of Bernanke's central recommendations was to place a higher weight on models with strong empirical relevance, ensuring that policy decisions are grounded in data-driven insights rather than overly rigid theoretical assumptions. Additionally, the review emphasized the importance of embracing a multiplicity of modeling approaches, advocating for a more diversified toolkit rather than reliance on a single dominant framework. This shift acknowledges the inherent uncertainty in macroeconomic forecasting and the benefits of combining structural, reduced-form, and data-driven models to enhance robustness. By integrating a broader range of models and prioritizing empirical accuracy, the Bank of England can improve real-time decision-making and better navigate economic shocks. This perspective aligns with broader debates in central banking, where the tension between theoretical consistency and empirical performance remains a critical issue in policy formulation.

2.3 Representative Agent vs. Heterogeneous Agent Models: Confronting Reality

The debate between Representative Agent (RA) models and Heterogeneous Agent New Keynesian (HANK) or Agent-Based Models (ABM) often centers on their alignment with both micro and macro data. The primary critique of RA models is their inconsistency with micro-level reality—there is no actual representative agent in microeconomic data. However, RA models maintain a strong connection with macroeconomic reality, as they start and end with macro-level data, ensuring internal consistency with national accounts.

HANK and ABM approaches attempt to resolve this micro inconsistency by introducing heterogeneous agents, thereby improving alignment with individual-level data. Yet, in doing so, they risk losing the macro-level coherence that RA models provided. This highlights a fundamental trade-off: while HANK and ABM correct RA's microeconomic limitations, they struggle with the macroeconomic consistency that RA models inherently preserve.

Both approaches, however, neglect a deeper issue: the aggregation problem arises after accounting for the measurement theory — the methodological choices involved in constructing national accounts, such as those governed by index number theory, which may be inconsistent with aggregation principles used in economic theory.[10] As Schlicht (1985) discusses in terms of open vs. closed aggregation, the way empirical macroeconomic aggregates are constructed does not necessarily align with the theoretical aggregation embedded in economic models as well as a conventional bottom-up perspective.

This disconnect reflects a broader issue often described as "*measurement without theory*", a debate dating back to Koopmans (1947), which questioned the reliance on empirical relationships without a guiding theoretical framework. While RA and HANK models both acknowledge the aggregation problem—that macro data must ultimately be reconstructed from micro foundations—they do so in different ways:

- RA models ensure empirical implementation remains anchored in macro aggregates.
- HANK models attempt to reconstruct macro-level aggregates using micro data redistributed from NA, implicitly assuming that these micro-level adjustments maintain theoretical consistency.[11]

As will be detailed, both models actually prioritize the micro theory and while overlooking aggregation in measurement when bridging the micro-macro gap.[12]

[10] However, this issue is well addressed in the index number literature, which typically compares the axiomatic and economic approaches behind the index numbers.
[11] Namely, it ensures consistency between data and model in solving the aggregation problem: individuals can be aggregated bottom-up to the macro on both fronts. Problematically, however, the adjusted micro data might not be real. We discuss such a critique in Section 6.
[12] According to Mirowski (1989, p.78), this methodological stance echoes Koopmans' position in the "*measurement without theory*" controversy: "Again, this is because neoclassical economics is an imitation of physics; he is simply incapable of entertaining the notion that there are alternative programs of legitimate economic research."

2.4 Discrepancies between theoretical and measured aggregation

While economic theory constructs aggregate variables based on idealized assumptions about individual behavior, national accounting relies on statistical and index number methodologies that may not conform to these theoretical structures.[13] This disconnect raises concerns about the validity of macroeconomic conclusions drawn from both micro-level estimates and measured aggregate data.

This section explores key evidence of this inconsistency, drawing from recent literature and empirical findings. Specifically, it examines how micro estimates used to answer macroeconomic questions depend on the structure of macro models, the challenges posed by the efforts close the micro-macro data gap, and the methodological differences between theoretical and measurement-based aggregation.

Firstly, recent research highlights how microeconomic estimates, when used to answer macroeconomic questions, must rely heavily on macroeconomic models rather than standing independently. This issue is explored in Orchard, Ramey and Wieland (2024), which examines the integration of micro-level estimates into macroeconomic models. To ensure the validity of both micro estimates and macro models, the authors propose constructing macro-level counterfactual scenarios and comparing them to historical benchmarks. They apply this approach to a case study of the 2008 U.S. tax rebates, revealing discrepancies between consumer spending behaviors predicted by integrating micro estimates with a macro aggregation model (counterfactual) and those inferred directly from aggregate NIPA consumption (actual), assuming no stimulus (see Figure 1). [14] Their analysis suggests that leading marginal propensity to consume (MPC) estimates may not fully capture the total consumption response, highlighting the need for a more nuanced integration of micro data into macroeconomic frameworks. Ramey (2019) stresses the necessity of returning to a macroeconomic context when interpreting empirical results for policy and practical relevance. Two central observations emerge:

[13] Beyond national accounting statisticians, as discussed below, the issue of incommensurable aggregation has been also been noted in the methodological literature. Janssen (2005, p. 51) reflects on this challenge, stating: "*These Bureaux collect statistics in a way that generally is independent of the specific wishes of economists…The problem I have with this view is that it is not clear why we should search for forms the aggregating functions have to satisfy if these complex aggregating functions are subsequently replaced by their published counterparts.*"

[14] This result simply updates the procedure in Sahm et al. (2012) to transform the micro estimates of MPC to spending volume induced by tax rebates and subtract it from actual spending to build the counterfactual. Considering equilibrium effects also makes the counterfactual scenario implausible, as suggested by Orchard et al. (2024).

- Micro estimates alone cannot answer macroeconomic questions: Their validity depends on the structure and assumptions of macroeconomic models, implying that there is no "applied micro free lunch" for macroeconomists.
- Differences between micro-based and macro-based results: Micro-level estimates, when aggregated through models, often yield answers that diverge from direct macro-level data.

What are the structure and assumptions of macroeconomic models when linking theoretical with empirical research? This discrepancy potentially highlights a fundamental aggregation issue: the framework and underlying principles of macroeconomic models may not incorporate the empirical structure used for measurement (e.g., index numbers behind the construction of NIPA consumption data). In Orchard et al. (2024), the bottom-up aggregation of expenditure from rebates could significantly diverge from NIPA's approach, particularly in how index numbers account for heterogeneity.[15]

The second piece of evidence comes from ongoing efforts to bridge the micro-macro data gap by redistributing aggregate data to individual households and firms. The implicit assumption is that the bottom-up aggregation of micro data should perfectly match its national accounts counterpart. However, the outcome from these efforts have potentially contradicted this assumption. The European Central Bank's (ECB) Distributional Wealth Accounts (DWA) project highlights such an example of aggregation inconsistencies. The DWA redistributes Financial Accounts (FA) in the Euro Area to individual households using data from the Household Finance and Consumption Survey (HFCS) to provide a more granular perspective on wealth distribution. Figure 2 shows that the linked data from DWA improves coverage ratios of financial aggregates, yet significant gaps remain—particularly in deposits, where a discrepancy of approximately 40% persists even after adjustments to account for the missing rich-tail distribution and the discrepancies on concepts and coverage between macro and micro datasets.

While it may be premature to conclude that all micro-measurement errors have been addressed, the persistent gaps in data redistribution projects suggest a deeper issue: national account aggregates used for macro policy analysis may not seamlessly translate to disaggregated data. Specifically, the macro relationship $F(\,)$ in (15), when explaining the relationship between two FA synthetic aggregates, or the aggregating function $G(\,)$ and $g_j(\,)$ in (12), when forming the nature aggregates from measurement, differ from

[15] Note that this divergence in aggregation methodology differs from the gap between micro-level consumption categorization and the BEA definition reconciled by Orchard et al. (2024).

those from theory. As a result, the aggregation problem fails from the outset. These discrepancies indicate that measurement-based aggregation does not always conform to theoretical or conventional expectations, which typically follow a bottom-up approach, potentially leading to distortions in macroeconomic analysis when relying on microfounded explanations.

Thirdly, the methodological divide between theory and measurement is widely recognized among national accounting statisticians. A prominent example is the well-known distinction between economic and axiomatic approaches in index number theory. CPI is often regarded as a bottom-up measure of price levels, yet its aggregation process — though justified by macro-level axioms — introduces potential distortions when viewed from a micro-theoretical perspective. The Boskin Commission report critically examines these biases and the conditions under which CPI can be considered a true bottom-up measure.

Another key debate concerns the interpretation of real GDP within the value-added framework—specifically, whether it can be viewed as the aggregate of firms' physical value added, a microfounded concept, under the widely used double-deflation approach. This issue has been extensively analyzed by David (1966), Sato (1976), and Meade (2010). We will further explore this topic in the following sections, linking it to recent quantitative assessments by Soyres, Gaillard, and Young (2023) for the U.S.

These examples illustrate the persistent gap between theoretical aggregation methods and those employed in macroeconomic measurement. Addressing this discrepancy requires a more explicit integration of measurement theories into macroeconomic modeling to ensure that empirical foundations align with theoretical frameworks. To achieve this, we formalize two distinct theoretical aggregation approaches—closed and open aggregation—in the next section. Specifically, the discussion focuses on the limitation of closed aggregation treated as a purely deductive enterprise.

3 Closed vs open aggregation

Schlicht (1985) distinguishes between two forms of aggregation—closed and open—a summary of which is provided in Appendix A. In this section, we demonstrate that the purely deductive structure of closed aggregation is incompatible with the pragmatic objectives of open aggregation. We then illustrate this discrepancy with a simple example using production function aggregation.

3.1 Closed aggregation as deductive exercise

This section argues that closed aggregation's deductive approach often fails empirically. Measurement acts as the gatekeeper for adequate solutions to the aggregation problem, prioritizing practical objectives like policy interpretation and actionable insights. As a result, some theoretically feasible aggregations—common in fields like physics—become operationally untenable in economics. Next we illustrate the evolution of aggregation in micro theory (closed aggregation) and aggregation in measurement (open aggregation) by a simple case of production function aggregation and discuss the implications in terms of causal policy analysis as well as the interpretation of micro-macro data gap.

In the Econometrica's debate on the aggregation problem, May (1946) and Pu (1946) took the stand of closed aggregation (Schlicht, 1985, p.71-79; Janssen, 2005, p.49). The aggregation of production functions, they discussed, exemplifies the deductive nature of closed aggregation. Consider the individual production functions and aggregating functions $\alpha(x)$

$$q_i = f_i(k_i, l_i),\ i = 1, \dots, n \tag{16}$$
$$K \coloneqq \sum_{i=1}^{n} k_i \text{ and } L \coloneqq \sum_{i=1}^{n} k_i \tag{17}$$

where q_i, k_i and l_i represent the firm i's output, capital and labor inputs, and K and L are the total capital and labor. Each firm maximizes the profits

$$\pi_i = pq_i - rk_i - wl_i \tag{18}$$

such that we obtain the typical restrictions

$$\frac{\partial q_i}{\partial l_i} = \frac{w}{p} \text{ and } \frac{\partial q_i}{\partial k_i} = \frac{r}{p},\ i = 1, \dots, n \tag{19}$$

Collecting together, there are $3n + 4$ variables $(q_1, \dots, q_n, k_1, \dots, k_n, l_1, \dots, l_n,\ K,\ L, \frac{w}{p}, \frac{r}{p})$ and $3n + 2$ equations as above. Under some regularity conditions, we apply the implicit function theorem to achieve

$$q_i = h_i^1(K,\ L), i = 1, \dots, n \tag{20}$$

Overall, the total production *function* is generally derivable as

$$Q \coloneqq \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} h_i^1(K,\ L) \coloneqq F(K,\ L) \tag{21}$$

Following the foundational discussion of aggregation (Section 2 and Appendix A), closed aggregation poses *no problem* (of consistency) as long as the open aggregation *can* select

the same macro restriction $F(\,)$ *a posteriori*—in this case, exactly as the aggregate function derived in (21).[16] However, as argued in Section 2.2, $F(\,)$ is imposed *a priori* by measurement through open aggregation. This creates a critical uncertainty: there is no guarantee that the inductively derived $F(\,)$ aligns with the deductively derived $F(\,)$ from closed aggregation. This mismatch—between open aggregation's inductive pragmatism and closed aggregation's deductive rigor—is the core problem.

The literature on the aggregation problem (e.g., Nataf, 1948; Gorman, 1953; Fisher, 1969) focuses on ensuring single macro aggregates to sidestep issues like multiple equilibria or competing macro functions. This preoccupation contrasts starkly with disciplines like physics, where many-to-many aggregation is intrinsic to explanations. Electromagnetic fields are a good example. The field at a point is the result of all charges and currents, so it's a many-to-many aggregation. Each charge contributes to the field, and the field affects other charges.

This divergence reflects economics' practical demands, not theoretical limitations per se. Scientific explanation in economics necessitates excluding many-to-many aggregations because economic measurement—unlike physics' descriptively oriented frameworks—imposes functional relationships to serve policy interpretation and operational guidance. Physics tolerates complexity for explanatory completeness (e.g., Maxwell's equations); economics prioritizes simplicity for policy action.[17] Functional models, such as the Phillips curve or Keynesian multiplier, provide straightforward levers for intervention, even if they oversimplify micro-level realities.[18] Economic measurement is inherently tied to policy needs, emphasizing real-world data and empirical validation, while economic theory often abstracts from these practical constraints.

3.2 An example: production function aggregation

In this section, we dissect how the chasm between ontological individualism and pragmatic constraints shapes the divide between microeconomic theory and macroeconomic measurement in aggregation approach. Then a simple model of aggregating production functions clarifies the differing approaches of microeconomic theory and macroeconomic measurement: closed versus open aggregation. The ensuing explanation discusses the implications for policy and microfounded interpretation.

[16] See the same argument in Schlicht (1985, p.72-73) and Janssen (2005, p.50).
[17] For instance, Maxwell's equations endured decades of initial incomprehension by physicists, even as they confounded contemporaries, yet ultimately became foundational to electromagnetism (Dyson, 2007).
[18] The Phillips curve originated as an empirical observation later adopted for policy, while the Keynesian multiplier emerged theoretically but gained traction through operational utility. Both reflect economics' pragmatic focus on policy instrument feasibility, even as theory shaped their interpretation and evolution.

3.2.1 Critique of closed aggregation in macro measurement

Before our illustration of the validity of open aggregation approach adopted by measurement, it can be helpful to present a defence against a possibly very strong entrenchment among modern macroeconomists: why is the closed aggregation not suitable for macro measurement as the bottom-up view of macro is so nature?

The justification for the microfounded macro economics is grounded in ontological individualism: macroeconomic phenomena are fundamentally reducible to their microeconomic constituents, with no independent *macro-level* existence. Guided by such a vision, bottoming up the equilibrium micro outcomes derived deductively (i.e. y_i obtained after considering the mutual compatibility of the *perfect* micro restrictions in (14); i.e., closed aggregation/micro theory), by the aggregating functions (12) identical to those in (15) (i.e. open aggregation/macro measurement), we should not expect any discrepancy between closed aggregation (14) and open aggregation (15). Here the situation of *perfect* micro restrictions presuppose we can model all dimensions of heterogeneity and theorize micro relationships with flawless precision about micro reality. National accounting, in this view, would merely reflect the mechanical summation of micro behaviors.

Leaving the assumption that we can exactly model the micro regularities alone, nevertheless, the information friction in practice challenge the macro measurement to adopt the same perspective. First, due to information asymmetry, measurement—constrained by incomplete data, unobserved heterogeneity, and computational limits—cannot access or model all micro-level details from the macro level, even if those restrictions perfectly describe individual behavior. Second, the regularities at the micro level are partitioned—the practical omniscience of the real world is mission impossible; they are not static and continuously evolving—a dynamic epistemology where knowledge is in a constant state of inquiry, testing, and revision. For example, it took centuries for the ideal gas law to be gradually reduced to kinetic theory through piecewise refinements. The example of a simple production function aggregation in the following Section 3.2.2 focuses on the distinction between two aggregation approaches prompted by information friction.

On the other hand, the explanatory goals differ: micro theory explains partitioned phenomena through *ceteris paribus* clauses, isolating causal mechanisms, whereas macro measurement is driven by practical objectives like prediction and policy interpretation. The former seeks causal fidelity; the latter, operational coherence. The case study on the gap between real GDP measured in national account and the microfounded physical value added in Section 4 addresses such a disparity.

Faced with these constraints, macro measurement resorts to *open aggregation* – aggregating observed outcomes first and then imposing macro restrictions (which may not be fully microfounded) – to obtain functional relationships that are operationally useful for policy purposes such as stabilization, forecasting, and intervention.

3.2.2 Reasoning of two aggregation approaches

Consider the same problem as discussed in Section 3.1 (May, 1946; Pu, 1946), two firms with the same production function

$$q_i = k_i^a + l_i^b,\ i = 1,\ 2 \tag{22}$$

The micro and macro variables are simply

$$x = \{q_i,\ k_i,\ l_i,\ i = 1,\ 2\} \text{ and } X = \{Q, K, L\} \tag{23}$$

Suppose we discover the first set of micro regularity – both maximize profit:

$$f^0(x) = \{x\ satisfying\ conditions\ (17) - (19)\} \tag{24}$$

In this context, the whole micro world has only one set of restrictions. Therefore, the resulting macro relationships via closed aggregation E_{f^0} and open aggregation $E_{F^0}^o$ are the same

$$E_{f^0} = E_{F^0}^o = \{X \in R^3\} \tag{25}$$

Since a and b are not fixed, profit maximization itself is not sufficient for aggregating these two individual production functions up to any distinct aggregate relationship. Figure 3 demonstrates this aggregate outcome in the three dimensional X space where closed and open aggregation overlap.[19] This case reveals that two aggregation approaches can coincide if information friction disappear.[20]

Next, we observe that the micro-level reality evolves according to a distinct regularity in the parameters a and b, ($a = 3,\ b \in \{2,3\}$), introduced alongside the principle of profit maximization. Consequently, the new set of micro restrictions becomes

$$f^1(x) = f^0(x) \cap \{a = 3,\ b = (2,3)\} \tag{26}$$

However, the national accounts do not have access to this parameter restriction and simply observe the macro outcome of closed aggregation. This lack of information creates

[19] For the visualization, we use many random discrete points to represent the aggregates can be any point in the cube.
[20] Here we reach at an omniscience state, as we do not have another partition of micro regularity to learn.

an asymmetry. Figure 4 demonstrates the macro relationships E_{f^1} generated by the closed aggregation (and observed by the open aggregation denoted as $E_{F^1}^O$)

$$E_{f^1} = E_{F^1}^O = \{X\ in\ the\ green\ and\ blue\ surfaces\} \tag{27}$$

where two firms aggregated to the blue surface when $a = 3$ and $b = 3$ and to the green surface when $a = 3$ and $b = 2$.

As the micro-level landscape continues to develop, a new regularity emerges alongside $f^1(x)$

$$f^2(x) = \{a = b\} \tag{28}$$

This condition introduces constant return to scale (CRS) to individual production function.[21] The national accounts are informed of this additional micro-level restriction. In the closed aggregation, we first impose these restrictions *jointly* in the micro level and then aggregate

$$E_F \coloneqq \alpha\left(E_{f^1} \cap E_{f^2}\right) = \{X\ in\ the\ blue\ surface\} \tag{29}$$

as depicted in Figure 5 the aggregate function corresponds to the single blue surface.[22] Here, $f^2(x)$ intersects with $f^1(x)$ at the micro level, resulting in $a = b = 3$.

Figure 6 illustrates the logical flow of open aggregation. In this approach, national accounts first aggregate partial equilibria separately to the macro level, denoted as $E_{F^1}^O = \alpha(E_{f^1})$ and $E_{F^2}^O = \alpha(E_{f^2})$. The intersection at the macro level is then established as

$$\begin{aligned} E_F^O &\coloneqq E_{F^1}^O \cap E_{F^2}^O = \alpha(E_{f^1}) \cap \alpha(E_{f^2}) \\ &= \{X\ in\ the\ green\ and\ blue\ surfaces\} \end{aligned} \tag{30}$$

As noted, open aggregation under the micro restrictions profit maximization and uninformed parameter restriction $E_{F^1}^O$ yields the macro relationships of *X* as portrayed by either the green or the blue surfaces. The macro outcome from the other open aggregation, $E_{F^2}^O = \alpha(E_{f^2})$, encompasses all points in this three-dimensional space of *X*.[23] Any point can be rationalized by aggregating individual firms with $a = b$ taking a specific real

[21] Strictly speaking, the production function with output transformed as $q_i^{\frac{1}{a}}$ has CRS.
[22] The functional relationship of *X* is derived from (17) − (19) with $a = b = 3$.
[23] We use many random discrete points to represent the aggregates can be any point in the cube.

number. Ultimately, the intersection of these two sets of macro outcome, E_F^O, coincides with the first set $E_{F^1}^O$.

3.2.3 Implications

The divergence between closed and open aggregation—rooted in ontological individualism versus pragmatic measurement— yields critical implications for policy analysis, micro-macro data alignment, and conceptual clarity in economics. Building on the production function case study, we interpret these results through two lenses:

1. **Policy Interpretation**

First, when analyzing policy-oriented time-series data, the distinction between closed and open aggregation shapes causal attribution. Suppose policymakers observe two aggregate production surfaces prior to $t-1$ (e.g., blue and green surfaces in Figure 4). At *t*, CRS are introduced as a micro-level regularity, collapsing observations to a single surface (blue) from $t+1$ on.

- **Micro perspective**: The shift is causally driven by CRS, as introducing this restriction in closed aggregation leads to eliminating incompatible outcomes.
- **Macro perspective**: Open aggregation, unaware of the parameter restriction, observes the collapse but cannot definitively link it to CRS through the causal reasoning. The policy narrative might instead cite exogenous shocks or measurement revisions.

This underscores the epistemic asymmetry between theory and measurement driven by information friction, adding complexity to policy feedback loops.

2. **The Micro-Macro Data Gap**

In a one-shot observation, the gap between synthetic aggregates (e.g., real GDP) and their microfounded counterparts (e.g., physical value added) hinges on whether micro restrictions are fulfilled:

- **No Gap (Blue Surface)**: If the observed aggregate *X* lies on the blue surface ($a=b=3$), real GDP (open aggregation) aligns with physical value added (closed aggregation). Here, micro restrictions (CRS, profit maximization) are satisfied, harmonizing measurement and theory.
- **Gap (Green Surface)**: If *X* resides on the green surface ($a=3$, $b=2$), real GDP diverges from physical value added. Open aggregation, lacking access to micro

parameter restrictions, cannot discern this mismatch, creating a latent measurement gap.[24]

The problem extends to concepts like capital income, where macro and micro terminology overlaps but measurement diverges.[25]

Collectively, the production function aggregation example vividly illustrates that while microeconomic theory envisions a seamless integration of individual behaviors through closed aggregation, practical constraints—such as information asymmetry and dynamic micro regularities— prevent macro measurement from relying on closed aggregation. This divergence is not merely technical; it embodies a profound epistemological gap that influences both policy interpretation and the alignment between microfounded constructs and observed macro aggregates. On the other hand, while theory provides foundational hypotheses, economic aggregation is ultimately disciplined by measurement's policy-driven imperatives—resolving empirical ambiguity to deliver operationally tenable explanations all in the macro level. The case study in the next section exemplifies this macro measurement priority.

4 Case study: Real Value Added vs. Physical Value Added

This section examines the disconnect between Real GDP (Real Value Added/RVA) in the national accounts and its microfounded counterpart, Physical Value Added (PVA).[26] By interrogating the index number theory underpinning RVA—a product of *open aggregation*—we reveal how macro-level explanatory priorities can diverge from a bottom-up perspective. We identify the conditions under which RVA approximates PVA, quantify their discrepancy when those conditions fail. Through this case study, we demonstrate why national accounting must rely on the open aggregation to meet its policy-driven imperatives.

4.1 Macro restrictions in practice

Independent from any concern of microfoundation, NA has its own macro level explanatory goals in order to match the practical policy framework. In the case of RVA, there are two fundamental objectives:

[24] This should not be interpreted as a critique of measurement. Instead, analysts aiming to microfound such aggregates must account for these potential discrepancies.
[25] Note capital income is actually part of value added.
[26] Consider a production function for individual firm, PVA is the sum of the factors contributing to the output quantity measure except the intermediate inputs, including capital, labor and techonology.

- *Consistency* btw income-expenditure and production sides of the real value added (RVA)
- Solving index number problem: separate the price and quantity components *synthetically* in the macro level

In the context of aggregation problem, NA simply codes these goals into macro restrictions $F()$ and the aggregating function accompanying the solution to the index number problem as $G()$ and $g_j\ ()$. These restrictions look all naturally valid from the macro perspective. But they are not necessary be deducted from micro theories, as interpreting RVA as PVA is not the prior goal behind.

Note such consistency is only maintained in the macro level, as will be shown, ensured by double-deflation—RVA is calculated as the residual between the value of real output and real intermediate inputs. The deflators for value added, output and intermediate inputs are constructed to only guarantee the deflated value added equal to the deflated output minus deflated intermediate inputs in the aggregate. Applying such macro deflators on each individual firm, the real value added as deflated is not necessarily correspond to the residual between the real output and real intermediate inputs in the micro level. Here is a didactic illustration: say, $x_i = y_i - w_i, i = 1,2$ and $\frac{x_1+x_2}{D_x} = \frac{y_1+y_2}{D_y} - \frac{w_1+w_2}{D_w}$, where x_i, y_i and w_i are respectively values (product of price and quantity) of value added, output and intermediate inputs, and D_x, D_y and D_w are the corresponding deflators. Maintaining the macro level consistency, we can still have $\frac{x_1}{D_x} = \frac{y_1}{D_y} - \frac{w_1}{D_w} - d$ and $\frac{x_2}{D_x} = \frac{y_2}{D_y} - \frac{w_2}{D_w} + d$, with $d \neq 0$, as deflated individual RVAs. Therefore, such a consistency restriction is not imposed on the micro level. On the other hand, note such residual identity is true for the nominal values on both macro and micro levels.

Similarly, price and quantity indices are constructed for the distinct explanation of macro level price and quantity changes. However, they do not necessarily represent any individual price and quantity movement—the index number problem is fundamentally about aggregating heterogeneous items. For instance, deflating any individual firm's output by the total output quantity deflator might not recover it to the base period since total quantity deflator is not necessarily same as the individual quantity ratio (index) between periods.

Following the previous example, assuming two firms are identical, $\frac{x_i}{D_x}$ and $\frac{y_i}{D_y} - \frac{w_i}{D_w}$ revert to their nominal values in the base period, as aggregate deflators equate to individual quantity ratios. As posited, these nominal values should consistently align, even at the micro level. This implies that, under homogeneity, RVA is indeed bottom-up

microfounded. But this is unrealistic.[27] Moreover, it is well-established that the index number problem is intricately linked to the first consistency requirement—the efficacy of disentangling price and quantity changes is defined by numerous axioms pertinent at the macro level.

Having established these two macro-level restrictions at an abstract level, we now operationalize them within the open aggregation framework and demonstrate their significance for policy-relevant modeling.

4.1.1 Consistency between income-expenditure and production accounts

The double-deflation procedure is widely recommended in the System of National Accounts (SNA), as well as by the United Nations (UN) and the European System of Accounts (ESA95), to ensure consistency between the income–expenditure and production sides of real value added (RVA).[28] In nominal terms, value added is derived as the residual between total output revenue and the cost of intermediate inputs, a relationship that directly reflects income generation. This inherent consistency in nominal value added serves as the basis for implementing a double-deflation approach.

However, translating this balance into real terms requires a parallel alignment in deflation methodologies. The total output and intermediate input figures have corresponding components in final demand and intermediate input spending, respectively. These components are deflated separately using distinct deflators—one for output and another for intermediate inputs—to accurately capture their changes over time.[29] The resulting dual adjustments are then combined to form the real value added in the production account, preserving the internal consistency of the national accounts. Alternatively, if the expenditure account is structured such that the same deflator (typically the output deflator) is applied to both output and input measures, a single deflation method can be used, as discussed by David (1966). Overall, whatever the deflation method—double, single, or otherwise—is ultimately adopted to satisfy the fundamental consistency

[27] This example does not entirely rule out bottom-up aggregation; it remains possible for the individual $\frac{x_i}{D_x}$ and $\frac{y_i}{D_y} - \frac{w_i}{D_w}$ to differ at the micro level while their aggregate sum is equal. We examine the conditions under which this scenario occurs in Section 4.1.2.

[28] See Meade (2010) and Sato (1976).

[29] This separate deflation approach also aligns with the operational and welfare considerations of the income account. The remuneration corresponding to these two expenditure categories typically accrues to distinct firms and population groups, necessitating disjoint data collection. Consequently, independent deflation is warranted to account for the differing preferences and consumption patterns of these income recipients.

restriction between the production and income–expenditure measures in national accounts.[30]

The coherence between income, expenditure, and production measures of real value added is a critical macroeconomic policy concern, as discrepancies among these measures can signal structural imbalances, measurement errors, or distortions in the economy. For example, during the post-2008 Global Financial Crisis, production-side value added in advanced economies (e.g., U.S. manufacturing) recovered more slowly than expenditure-side consumption, which was driven by debt-financed household spending. Although double-deflation is theoretically designed to ensure consistency between production and expenditure accounts, practical discrepancies persist due to measurement challenges in key components such as deflators, sectoral coverage, and data collection. These inconsistencies hold diagnostic value: they not only guide policymakers toward structural interventions but also expose gaps in measurement frameworks that demand refinement.

Several representative instances underscore these challenges. First, falling manufacturing output prices (driven by automation and efficiency gains) alongside sharply rising input prices (e.g., energy costs) distort real value added when outdated or mismatched deflators are applied. Second, post-2008 U.S. households leveraged debt to purchase imported goods—a dynamic that inflated expenditure-side growth while domestic manufacturing stagnated, highlighting a misalignment between consumption-driven demand and domestic productive capacity. Third, production-side data (e.g., industrial output) is often revised with a lag, whereas expenditure-side metrics (e.g., retail sales) are timelier but prone to initial inaccuracies, creating transient inconsistencies in real-time policy analysis.

Crucially, such inconsistencies are not mere statistical artifacts; they serve as critical signals for policy action. During the Financial Crisis, central banks maintained low interest rates to stimulate production-side investment, while governments introduced job-retention schemes to stabilize incomes and incentivized reallocation of productive capacity (e.g., from real estate to technology sectors). Simultaneously, these discrepancies exposed the urgency of addressing measurement limitations, prompting methodological upgrades such as the adoption of chain-weighted indices, improved tracking of informal and digital activity, harmonized data revisions to reduce timeliness gaps, and granular regional-sectoral metrics to map debt and import exposure.

[30] Meade (2010) identifies extrapolation as another key method. As will be highlighted below, Soyres et al (2023) further demonstrate how bottom-up PVA can be calculated (if microfounded interpretation of RVA is accepted).

Thus, verifying accounting identities in value-added measurement performs a dual function: they reveal economic vulnerabilities requiring immediate policy responses while sounding an alarm for systemic improvements in data infrastructure. By reconciling these gaps—through refined deflators, expanded sectoral coverage, and synchronized revisions—policymakers can better align macroeconomic narratives with underlying realities, ensuring interventions are both timely and grounded in robust empirical foundations.

4.1.2 Solving the index number problem

The index number problem is fundamentally a challenge of synthetic construction—an effort to reduce multidimensional, heterogeneous microeconomic changes (e.g., prices and quantities of millions of goods) into singular, scalar measures of aggregate price (P) and quantity (Q) that are inherently unobservable. Concepts like "real value added" or the "general price level" are epistemic tools: they exist only as theoretical abstractions, derived from aggregating disparate elements into a cohesive narrative about the economy. For instance, while individual prices and outputs are observable, their collective behavior—such as inflation or real GDP growth—is not directly measurable. Instead, these macro-level phenomena must be synthesized through index-number methodologies, which impose structure on chaos by reconciling vector-based data (e.g., of individual price relatives and quantity relatives) into scalar indices.

This synthesis relies on two complementary pillars. The economic approach grounds the separation in microeconomic theory, treating P and Q as dual reflections of optimizing behavior. For example, when analyzing changes in the micro components, one can examine the ratio of cost functions while holding output constant to derive an input price index. Similarly, real output growth is inferred from revenue functions, contingent on hypothetical producer responses to price shifts. These indices are synthetic because they represent counterfactuals—what quantities would have been if prices changed, holding technology or preferences constant.

The axiomatic approach, meanwhile, imposes mathematical rules to ensure the synthetic separation adheres to intuitive principles, even in the absence of observable benchmarks. Axioms like proportionality ("if all prices double, P doubles") or factor reversal ($P \times Q$=Value Ratio) act as guardrails, ensuring indices behave logically despite their unobservable targets. However, these rules are ultimately conventions, designed to balance practical explanatory efficiency and computational convenience against the unnecessary burden of more complex, non-scalar representations. Scalar indices like CPI or GDP growth simplify the interpretation of multidimensional data, enabling policymakers and the public to grasp trends like inflation or economic performance

without grappling with the full complexity of underlying microeconomic dynamics. While alternative frameworks (e.g., vector-based indices or disaggregated measures) might offer greater precision, they often impose excessive cognitive and operational costs, making scalar indices the preferred tool for communication and decision-making despite their inherent simplifications.

The artificial yet systematic nature of this separation is indispensable for policymaking. Consider three dimensions:

- Interpretation: Central banks targeting inflation rely on synthetic price indices (e.g., CPI) to interpret diffuse price movements as a single "inflation rate" even though no entity directly experiences the "average" inflation of the basket. This synthesis is not merely a technical exercise but fulfil their implicit mandate to provide a common reference point for the diverse stakeholders, balancing competing needs for accuracy, transparency, and stability in economic decision-making. [31] Similarly, real GDP growth—a synthetic quantity index—frames macroeconomic performance as a scalar, enabling comparisons across time or nations.

- Action: Policy tools like interest rates or fiscal stimuli are calibrated to these synthetic metrics. For example, a central bank raising rates to curb inflation implicitly trusts that the CPI's axiomatic and economic foundations meaningfully isolate demand-driven price pressures from supply-side shocks.

- Monitoring: Synthetic indices simplify complex systems into tractable indicators. The European Central Bank's harmonized inflation index (HICP) or the UN's Sustainable Development Goals (SDGs) use aggregated indices to monitor progress, despite the inherent loss of granularity.

Ultimately, index-number synthesis is a balancing act: it must simplify reality enough to inform decisions while retaining enough theoretical rigor to avoid distortion. As economies grow more complex, this balance becomes ever more critical—and precarious. Hybrid frameworks (e.g., chain-linked Fisher indices) that blend economic theory with axiomatic flexibility exemplify the ongoing effort to refine these synthetic tools, ensuring

[31] The heterogeneous policy demand of inflation metrics includes fiscal compensation, social welfare adjustment, wage bargaining, financial markets and investors, business planning, informed spending and saving decisions, cross-country policy coordination and so on.

they remain fit for purpose in a world where unobservables like "real growth" or "price stability" shape real-world policy, markets, and livelihoods.

4.2 Conditions to close the gap

We compare the RVA measured in national account with the PVA grounded in micro theory, and explore the conditions to equate the two.[32] Following Sato (1976) and Sims (1969), we employ the Divisia index to derive the measurement framework of RVA, as it provides the most generalizable approach for connecting axiomatic and economic index number theories with minimal assumptions.

To begin, we outline the construction of the Divisia index.[33] Suppose there exist vector-valued functions $p(t) \equiv (p_1^t, \dots, p_N^t)$ and $x(t) \equiv (x_1^t, \dots, x_N^t)$ representing prices and quantities of N commodities at time t. We can derive

$$\frac{dln(p(\tau) \cdot x(\tau))}{d\tau} = \sum_{n=1}^{N} s_n(\tau) \frac{dlnx_n(\tau)}{d\tau} + \sum_{n=1}^{N} s_n(\tau) \frac{dlnp_n(\tau)}{d\tau} \tag{31}$$

where $s_n(t) \equiv \frac{p_n(t)x_n(t)}{p(t)x(t)}$, $(n = 1, \dots, N)$ denotes the value share of commodity *n*. From this, we define the Divisia price and quantity indices as

$$\frac{dP^{Div}}{P^{Div}} \equiv \sum_{n=1}^{N} s_n(t) \frac{dp_n}{p_n} \text{ and } \frac{dX^{Div}}{x^{Div}} \equiv \sum_{n=1}^{N} s_n(t) \frac{dx_n}{x_n}, \tag{32}$$

which tackles the index number problem.[34]

Next, we apply the Divisia index to decompose value added at the industry level. Starting from the nominal value-added identity for an industry

$$vV = qQ - mM \tag{33}$$

where V, Q and M are the quantities of value added, total output and intermediate inputs, respectively, and v, q and m represent their corresponding prices; v and V are scalars, others are to be regarded as vectors. Treating qQ and $-mM$ as additive subcomponents of vV, we apply the Divisia formulation again to derive

$$\frac{dV}{V} = (1 + \mu) \frac{dQ}{Q} - \mu \frac{dM}{M} \tag{34}$$

[32] This is a widely examined topic in the measurement literature. See Meade (2010) for an overview and Sato (1976) for an in-depth analysis of the underlying principles.
[33] See Balk (2005) for further details.
[34] The index number problem is addressed in the sense that the change (time derivative) of the value (product of price and quantity) with respect to the base value (i.e., lhs of equation (31)) is decomposed into the price and quantity changes.

$$\frac{dv}{v} = (1+\mu)\frac{dq}{q} - \mu\frac{dm}{m} \tag{35}$$

where $\mu = mM/vV$, $\frac{dQ}{Q}$ is the Divisa quantity index of individual products, and analogous interpretations apply to other terms.

To align RVA with PVA, RVA must equal the aggregation of PVA across firms within an industry. Consider an industry-level production function

$$Q = F(X, M) \tag{36}$$

where Q, X and M represent scalar output, vectors of primary factor inputs and intermediate inputs, respectively. Differentiating (36) yields

$$\frac{dQ}{Q} = \frac{F_X X}{F}\frac{dX}{X} + \frac{F_M M}{F}\frac{dM}{M} \tag{37}$$

where $\frac{dX}{X}$ and $\frac{dM}{M}$ are weighted sums of $\frac{dX_i}{X_i}$ and $\frac{dM_i}{M_i}$ whose weight are $\frac{F_{X,i}X_i}{\sum F_{X,k}X_k}$ and $\frac{F_{M,i}M_i}{\sum F_{M,k}M_k}$, etc.

Combining (34) and (37), it follows that

$$\frac{dV}{V} = \underbrace{(1+\mu)\frac{dQ}{Q} - \mu\frac{dM}{M}}_{NA\ approach\ (Divisia\ index)} = \underbrace{\frac{F_X X}{Q - \frac{m}{q}M}\frac{dX}{X} + \frac{qF_M M - mM}{vV}\frac{dM}{M}}_{closed\ aggregation\ (general)} \tag{38}$$

Under CRS and profit maximization, the quantity index of value added $\frac{dV}{V}$ coincides with $\frac{dX}{X}$, the relative PVA change derived from micro theory.[35] In this sense, RVA measured in NA can aligns with the aggregated PVA only if CRS and profit maximization under competitive market hold in the micro reality. When CRS holds, if marginal revenue deviates from marginal cost, $F_M \neq \frac{m}{q}$, there are distortions such as markups, taxes or other factors. The multiplier in the first term will not be unity and that in the second term will be nonzero. The wedge between RVA and PVA arises. In the presence of increasing (decreasing) returns to scale, even when $F_M = \frac{m}{q}$, the multiplier in the first term is not unity. Sato (1976) characterizes this layer of the wedge as the contribution arising from intangible primary factors of production, distinct from distortions.

[35] $F_M = \frac{m}{q}$ is the first-order condition of profit maximization, and this, with CRS, $Q - \frac{m}{q}M = Q - F_M M = F_X X$.

4.3 A quantitative evaluation

Soyres et al. (2023), focusing on the distortions, quantifies significant discrepancies between RVA and PVA.[36] The analysis reveals that sectoral growth and productivity metrics are systematically distorted by markups and intermediate input dynamics, with profound implications for economic policy and research.

The study highlights stark sectoral biases in growth measurement. In the Finance sector, RVA overstates cumulative growth by 20 percentage points (1999–2021) due to high markups and profit generation from intermediate inputs, which inflate reported value added without corresponding gains in physical output. Conversely, in Manufacturing, RVA understates growth by 21pp, as heavy reliance on intermediate inputs (70% of costs) obscures true productivity improvements. Agriculture exhibits the largest volatility gap, with PVA growth 76% more volatile than RVA, reflecting unaccounted fluctuations in input usage. At the aggregate level, these biases partially offset, masking sectoral distortions: cumulative U.S. RVA growth trails PVA by 1.24pp over the same period.

Productivity metrics are similarly skewed. The Solow Residual—a widely used productivity measure—overestimates growth in Finance by 15pp by conflating markup-driven profits with efficiency gains. In contrast, Manufacturing's productivity gains are underestimated by 15pp, as intermediate input costs dilute measured efficiency. These distortions arise because RVA mechanically incorporates profits from intermediate inputs, even when physical output remains unchanged.

Although the whole system of NA can be considered an open aggregation framework, the measurement theory behind is also influenced by the closed aggregation. Appendix B illustrates that the constrast between economic vs. axiomatic approaches in the index number theory actually showcases the coexisting influences by two aggregations and their limitations. At the same time, we also rationalize those natural-intuitive principles imposed by the axiomatic approach and demonstrates a convergence of both aggregations on the Divisia index, which can be useful in resolving the aggregation problem.

5 Fundamental barrier of aggregation in measurement

Aggregation in measurement faces inherent challenges rooted in the tension between theoretical ideals and empirical feasibility. At its core, macro-level measurement

[36] Their production function is a Cobb-Douglas one. We also observe that their formulation of the wedge between RVA and PVA in equation (18) is identical to our formulation in equation (38), assuming CRS, a result that follows trivially.

demands dimensionality reduction, a process fraught with impossibility. Early work by Ijiri (1971) and Demski (1973) underscores this dilemma: while neoclassical theory often sidesteps heterogeneity by assuming homogeneity—treating apples and computers as interchangeable units—real-world measurement cannot ignore the dimensionality problem. Quantities lacking a common unit resist meaningful summation, exposing the fragility of aggregation in practice.

This challenge deepens when considering the implications of Sonnenschein-Mantel-Debreu (SMD) theorem.[37] Theorists note that even when micro-level models are restricted by assumptions ensuring well-tested micro regularites, these constraints may cancel out during aggregation. Kirman (1989) and Janssen (2005) illustrate this paradox: for instance, applying maximum likelihood estimation (MLE) to separate micro-data subsets may yield excellent fits within each subset, yet when these subsets are aggregated, the overall coherence—or joint fit—can collapse. Characterizing macro measurement as open aggregation provides an alternative interpretation: the issue lies in the distinction between closed aggregation (even assuming interdependencies like feedback mechanisms) and *open aggregation*, which operates without access to joint micro-level restrictions. This creates the wedge between the empirical validity of the micro restrictions and their aggregate implications.[38]

Empiricists face parallel hurdles, shared by both data users and macro measurement practitioners. Synthesizing micro-data into macro-measurements under general equilibrium (GE) frameworks, as Hansen and Heckman (1996) emphasize, is fraught with impracticality. Micro studies often rely on partial equilibrium models, ignoring cross-market interactions or collapsing them into constants—a convenience that masks complexity. Shoven and Whalley (1992) highlight how benchmark data fail to uniquely identify parameters, relying instead on ad hoc elasticity values drawn from fragmented literature. This patchwork approach clashes with macro-measurement's need for universality, where aggregation demands unconditional knowledge of all parameters—a task Schlicht (1985) likens to *no isolation in aggregation*.

[37] The Sonnenschein–Mantel–Debreu (SMD) theorem, developed in the 1970s by economists Hugo Sonnenschein, Rolf Mantel, and Gérard Debreu, addresses the characteristics of excess demand functions within general equilibrium theory. The theorem demonstrates that even if individual consumers exhibit rational, utility-maximizing behavior, the aggregate market demand does not necessarily inherit well-behaved properties such as uniqueness or stability of equilibrium. Specifically, the market's excess demand function can exhibit almost any shape, provided it satisfies continuity, homogeneity of degree zero, and adheres to Walras's law.

[38] Kirman (1989) argues that interactions between micro and macro layers (e.g., feedback loops) could mitigate aggregation paradoxes, but such dynamics remain absent in open aggregation.

Closed aggregation, in theory, promises systematic coherence by requiring complete micro-level information. Yet this ideal crumbles under scrutiny. Micro identification is inherently *conditional*, depending on fixing external parameters—*no free lunch in identification*.[39] This mirrors the arduous challenges highlighted by Hansen and Heckman (1996): in our earlier example (Section 3.2.2), micro identification is achievable only by knowing $a = 3$, a parameter that remains uninformed at the macro level. Macro-measurement, by contrast, demands *unconditional* integration, creating a circular dependency where identifying one parameter hinges on others. Practically, this necessitates an infinite search akin to tracking quantum Bose particles, a Sisyphean task.

The resolution lies in embracing open aggregation. By abandoning the quest for closed-system completeness, open aggregation constructs macro variables with minimal assumptions, prioritizing feasibility over theoretical purity. While it sacrifices the elegance of joint micro-level fitting, it acknowledges the sparse, contradictory reality of empirical data. This pragmatic approach mirrors the struggles of both theorists and empiricists, bridging the gap between micro-level fragmentation and macro-level coherence. In essence, the fundamental barrier of aggregation is not merely technical but epistemological: measurement must navigate a world where perfect synthesis is impossible.

6 Redistributing NA and Representative Agent thinking

Redistributing National Accounts (NA) data addresses a fundamental aggregation problem: the gap between micro theory–based aggregates derived from micro data and the aggregates produced by NA methodologies. In practice, "redistributing" NA data is an attempt to close this gap, yet it exposes inherent methodological differences—the data gap persists after many reconciliation steps from the micro side as discussed in Section 2.4. Merely "redistributing" the NA sum presupposes that the fundamental differences between closed and open aggregation have already been reconciled—a presumption that is, in fact, erroneous. This section examines challenges extending beyond the fundamental gap between closed and open aggregation, alongside case studies demonstrating an unintentional reversion to representative agent thinking in redistributing FA.

6.1 Residual accounting and Barnett Critique

Beyond the fundamental gap between open and closed aggregation, two consequential arguments—arising from this gap and its interplay with the inherent structure and

[39] Essentially, the expression underscores that there is no identification method that is free from the cost of untestable or exogenous assumptions.

development of NA—potentially undermine the practice of "redistributing" NA. First, the household sector, treated as a residual in NA, can inherit the fundamental gap from other sectors (eg RVA vs PVA in corporate sector). In national accounts, claiming that the household sector is treated as a residual means that its economic figures—such as income, savings, and expenditure—are not measured directly. Instead, they are derived by subtracting the directly measured contributions of other sectors (like government, corporations, and the financial sector) from the overall national totals. This approach implies that any measurement errors or omissions in the other sectors automatically fall onto the household sector, making its data potentially less precise and more subject to aggregation issues.

Second, echoing the Barnett Critique, macroeconomic measurement inherently embodies a dual role: it is both shaped by and constrains theoretical coherence. The practice of "redistributing" NA, however, embodies a contradiction in this relationship—it attempts to impose microfoundational consistency (closed aggregation) on aggregates rooted in open-aggregation frameworks (e.g., GDP, sectoral accounts), thereby perpetuating their unresolved tensions

While the household sector's residual status exposes a direct cause of NA's aggregation gap, our focus here is on a subtler systemic tension—the Barnett Critique and on the currently concentrated efforts of redistributing the FA aggregates in the household sector. Developed by William A. Barnett, this Critique contends that simple-sum monetary aggregates/FA (e.g., M1, M2) are fundamentally flawed for policy analysis.[40] Simple-sum aggregation treats all monetary assets as equally substitutable (e.g., cash and bonds within monetary aggregates), violating microeconomic principles of marginal utility and substitutability. By ignoring differences in liquidity, risk, and returns, they distort the measurement of the "money stock" relevant for macroeconomic policy (e.g., inflation targeting).

Barnett's critique targets reduced-form policy models that rely on macro-level statistical regularities (e.g., correlations between money supply and GDP). These models operate under open aggregation, where aggregates like GDP or household saving are not necessarily constructed from microeconomic theory (see Section 2.2). Barnett advocates replacing simple-sum aggregates with Divisia measures, which weight assets by their marginal contributions to liquidity services—a method grounded in microeconomic theory. Divisia aggregates resolve the critique by harmonizing micro theory (closed aggregation) with index-number axioms (open aggregation), as shown in Appendix B.

[40] While FA incorporate heterogeneity through sectoral disaggregation and instrument-specific valuation, within sector and instrument simple-sum still prevails.

This convergence allows Divisia to serve as a bridge between microfoundations and macro-measurement, making it suitable for policy models reliant on NA data. Indeed, Chrystal and MacDonald (1994)—who coined the term "Barnett Critique"—demonstrated that Divisia aggregates exhibit stronger causal relationships with key economic variables compared to simple-sum aggregates.

Paradoxically, simple-sum aggregates could coexist with microfounded structural models (closed aggregation) if the model´s microfoundations explicitly capture the micro behavior and heterogeneity. [41] This rationale appears to justify the practice of "redistributing" FA as it is just a dual of simple-sum. Yet, policy analysis cannot be confined solely to FA only; it must link them to broader NA aggregates like GDP and productivity, which remain measured by open aggregation. On the other hand, "redistributing" NA appears valid if policy analysis adopts a fully microfounded structural model covering the entire economy. The challenge, however, is that aggregates should not be generated using distinct aggregation approaches within a single analysis. The solution may lie in not only "redistributing" FA but also addressing other areas of NA. But other NA areas are confronted with the fundamental gap between aggregation approaches (e.g., RVA vs. PVA). As a result, the problem becomes reconciling micro data from the FA—which has been "redistributed"—with data from other accounts that have not, or that rely on categorically different aggregation concepts (e.g., markup versus distributed capital income/physical value added). Ultimately, the Barnett Critique underscores a systemic tension: while Divisia-like reforms can align specific aggregates with theory, holistic reconciliation requires overhauling NA's entire measurement framework—a task complicated by entrenched institutional practices and conceptual divergences across sectors.

6.2 Two case studies: DFA and DWA

A case in point is the Distributional Financial Account (DFA) from the Federal Reserve Bank for the United States. Here, DFA interpolates distributional variables (for instance, the p90 of wealth) by relying on historical time-series relationships with quarterly NA aggregates (Batty et al., 2020). However, the structure linking existing micro and macro data remains opaque. In practice, nowcasting carries forward an intrinsically macro relationship—for example, a built-in macro restriction where aggregate demand (Y) is proportional to aggregate income (X) ($Y = aX$). Consider a simplified illustration: with only two identical households, the DFA's approach effectively reverses the aggregating function (yielding $y = Y/2$ and $x = X/2$ at the micro level), yet when these figures are

[41] In other words, any aggregating function—not just the simple sum—can be used, provided it aligns with the derived aggregate relationships.

reaggregated bottom-up, the macro estimation recovers $Y = aX$. This "celebration" of a microfounded explanation is tempered by the persistent need, which is still missing, to verify whether the redistributed micro data truly reflects micro-level realities or whether a more microfounded nowcasting methodology might be required. Notably, current interpolation practices hinge on open aggregation restrictions from NA—which may diverge from a bottom-up, closed aggregation framework. For instance, part of the growth in real income interpreted as value added might not translate into distributed income, with some of the increase maintained as retained earnings and potentially redirected to investment rather than consumption.[42]

In the Euro Zone, the Distributional Wealth Account (DWA) offers another illustration of a common practice among NA redistribution projects in closing the aggregation gap: effectively scaling up in the last step.[43] This method assumes homogeneity in the remaining gap (e.g. equal underreporting behavoir) to achieve exact aggregation; however, the reasoning behind is inherently regressive to the RA framework.[44] When deductions or reductions are built upon a reduced, macro-level theory, the explanatory power of microfoundation can diminish (Sarkar, 2015). An alternative strategy would involve explicitly studying the assumptions that link micro and macro data to characterize and quantify the resulting bias—a point made by David (1966): "*the power of making an explicit set of assumptions is that it is possible to determine the nature (and, in principle, the magnitude) of the bias in the measure of net output caused by any specific instance of their violation.*"

This discussion resonates with earlier considerations (Sections 2.3 and 2.4) on the aggregation problem. In Section 2.3, the debate between Representative Agent and Heterogeneous Agent models underscored that while RA models maintain macro coherence by starting and ending with aggregate data, they neglect micro-level heterogeneity. Conversely, models that attempt to reconstruct macro aggregates from redistributed NA micro data—although offering a more microfounded approach—often struggle to align with the measurement conventions of national accounts. Section 2.4

[42] Batty et al. (2021) explicitly acknowledge the methodological ambiguity in their analysis: it remains unclear whether their model forecasts a microfounded distributional explanation of macroeconomic trends (e.g., wealth inequality driven by heterogeneous household saving behavior) or identifies a purely macro-level statistical relationship (e.g., a statistical artifact correlated with macroeconomic shifts lacking clear microfoundations, for example, service-sector productivity shocks). Empirical evidence shows that macro relationships observed from NA, constructed via open aggregation, often deviate sharply from historical norms in times of significant external turbulence.

[43] The previous steps include the micro-measurement correction: adjustment of non-response and population coverage, micor-macro conceptual alignment as well as tail extrapolation.

[44] In the empirical narrative, RA simply scales down the aggregate to a mean person, applies the rational choice micro theory and finally scales up to return back to the national level

further elaborated on how the discrepancies between theoretical aggregation and measurement-based aggregation can lead to biases in macroeconomic analysis. The tension between these approaches highlights that the aggregation problem is not simply a technical challenge of data redistribution, but a deeper issue arising from the mismatch between empirical measurement methods and theoretical constructs.

Together, these insights emphasize that any effort to "redistribute" NA data must grapple with the fundamental gap between the micro foundations of economic behavior and the macro aggregates generated by measurement conventions. Otherwise, the illusion of '*measurement without theory*'—a glaring flaw of RA—remains. Only by explicitly linking micro and macro assumptions can we hope to quantify—and eventually bridge—this bias in order to construct more robust, empirically grounded macroeconomic models.

7 Tentative solutions

In addressing the fundamental gaps between microfounded theory and macroeconomic measurement, we propose several tentative solutions.[45] These solutions aim to reconcile the discrepancies by reintroducing a linkage between bottom-up micro models and the aggregate statistics observed in national accounts.

7.1 Satellite accounts

One promising approach is the development of satellite accounts, as discussed by Bos (2009, p.233-259). Satellite accounts allow us to recover the micro-level variable x by relaxing or "dropping" the micro restrictions, e.g., E_F in equation (29), inherent in the original micro measurement. Similarly, they recover the aggregate X by forgoing the macro restrictions, e.g., E_F^O in equation (30), imposed during national accounting. While this flexibility allows analysts to bypass rigid measurement assumptions, a key limitation is that Satellite Accounts may produce aggregates inconsistent with observed NA data, which are often central to policy and economic analysis.

A notable pro *and* con of Satellite Accounts lies in their handling of aggregation. Analysts are free to impose closed aggregation restrictions to model micro-macro feedback, but this choice hinges on the specific phenomena under study. For instance, whether non-NA feedback mechanisms (e.g., behavioral responses to policy changes) are relevant depends on the explanatory goals. This decouples measurement from modeling, freeing

[45] Barnett's Divisian approach to synthesize the microfoundation into index number is also a solution to facilitate satisfying reduced form analysis; however, its use in a microfounded context is still unclear, which is our focus here.

researchers from rigid micro/macro constraints that traditionally bind NA frameworks (Bos, 2009).

7.2 Bottom-up Approaches in the Context of Productivity Analysis

Another line of inquiry involves bottom-up approaches prevailing in productivity analysis (Balk, 2016). The goal here is to explain aggregate productivity indices using detailed micro data and models. In this framework, micro data are collected and used to construct models—potentially incorporating neoclassical assumptions—to fit individual firm-level productivity measures. A major pitfall in this approach is the arbitrariness of weighting schemes during aggregation (Loecker and Konings, 2006; footnote 5). The measurement of aggregate productivity is governed by the open aggregation macro restrictions, which do not necessarily discipline the connection with micro level productivity. Basu and Fernald (2002) note the choice of weights can be manipulated to produce a desired aggregate outcome.[46] Mostly these weights may not reflect the true underlying interaction of micro productivity. Moreover, consistency in aggregation can be violated, as Reich (2016) argues. (In)consistency in aggregation means that productivity (index) aggregated up to any subaggregate level (e.g. industry and sector) and then aggregated to the national level may not be the same as the one which is aggregated directly from bottom-up (i.e., firm level). This is simply an aggregation problem for the measurement.[47] Therefore, while bottom-up approaches promise a microfounded explanation, their success depends critically on the rigorous and meaningful selection of weights during aggregation.

7.3 Top-down Approaches in the Context of Productivity Analysis

Complementing the bottom-up methods are top-down approaches, which decompose aggregate productivity indices to uncover the underlying weights and structure inherent in national accounts. This approach respects the NA aggregation methodology—as detailed by Balk (2016, 2018)—which is based on established accounting identities as the primary micro-macro linking principle. For example, the nominal value added index and the input cost index are constructed in NA such that the aggregate is a sum of the corresponding subaggregate components. Top-down methods do not necessarily resort to neoclassical micro assumptions; rather, they focus on maintaining the measurement practices by using index number techniques and searching the weights to address the aggregation problem unresolved in the bottom-up approach. These methods include: (a) adopting index numbers instead of theory driven price and quantity aggregates; (b)

[46] . For instance, for a set of values such as {1,2,3}, there exist countless weighting schemes for them to add up to a specific average (e.g., 1.5).
[47] Appendix C offers a concise overview of this matter.

recognizing that NA-specific aggregating functions are often complex and non-linear; (c) netting out intra-unit transfers within firms or industries to avoid double-counting; and (d) adjusting for firm turnover to isolate organic productivity growth from compositional changes. By decomposing the aggregate productivity index, top-down approaches offer a reverse-engineering perspective that reveals whether the observed macro measures are consistent with the underlying micro dynamics.

Both satellite accounts and the combination of bottom-up and top-down approaches provide promising, albeit tentative, solutions to the persistent micro–macro gap. By reintroducing a rigorous linkage between theory and empirics, these solutions seek to overcome the paradoxical regression to representative agent thinking and improve the fidelity of aggregate economic analysis.

8 Conclusion

Klein (1946b) delineates two fundamentally distinct paths for addressing aggregation:

"There are at least two essentially different approaches to the problem of aggregation. We may accept the traditional theories of microeconomics and the commonly used aggregates such as the Federal Reserve Board production index, the Bureau of Labor Statistics price index, the Commerce Department national-income data, etc., and try to determine the structure of a macroeconomic system that is implied by these two sets of information. Such a system may be very complicated or may not even exist, but if we could construct it, we could hope to make forecasts of these well-known aggregates. Alternatively we could proceed differently by assuming the theories of micro- and macro-economics in advance and then discovering what aggregates are consistent with these assumptions. In this case, we cannot know in advance the form of the aggregates but must accept those forms which satisfy a mathematical requirement. In the latter approach we cannot in general claim to have models that help us to predict the Federal Reserve Board index or the Bureau of Labor Statistics price index. Instead we obtain models that attempt to forecast the particular aggregates that satisfy the criteria assumed. It will often be true that in practice the correlation between these aggregates obtained by the second method and the published indexes will be so high that one set can be substituted for the other."

Klein's argument hinges on an assumption of *high substitutability* between the aggregates derived from these approaches. Yet this premise—that methodological divergence is ultimately bridged by empirical correlation—faces mounting skepticism as our study shows. On the other hand, the rise of computational power, granular microdata, and sophisticated modeling techniques has transformed the "second approach" (theory-first

aggregation) into a far more intricate and computationally advanced framework, eclipsing its mid-20th-century limitations. Can we still defend Klein's portrayal of the "first approach" (measurement-first modeling) as inherently complex or impractical relative to the second, given that modern tools now enable unprecedented precision in reconciling microfoundations with measurement constructs?

The chasm between microeconomic theory and National Accounts (NA) measurement is not merely technical but foundational—a divide rooted in conflicting priorities. While economists prioritize microfoundational consistency, NA statisticians navigate the pragmatic constraints of synthesizing heterogeneous data into policy-relevant aggregates. Bridging this gap demands more than methodological tweaks; it requires a paradigm shift toward collaborative engagement with NA practitioners.

Our analysis reveals that efforts to "redistribute" NA frameworks to align with micro-theoretical ideals risk regressing into a paradoxical revival of representative agent (RA) thinking. By imposing closed aggregation logic on inherently open systems, such attempts replicate the very oversimplifications they seek to overcome. Instead, progress lies in rationalizing the NA statistician's perspective: their focus on open aggregation, macro coherence, and operational feasibility is not a flaw but a necessity. Recognizing this eases communication with National Statistical Offices (NSOs), fostering mutual trust in data access, reconciliation of micro-macro discrepancies, and nuanced policy interpretation.

Central Banks emerge as pivotal actors in this reconciliation. Their mandate—interpreting macroeconomic trends and guiding policy—depends critically on understanding aggregation. Misreading NA aggregates (e.g., conflating open aggregation outputs with micro-founded constructs) can distort policy responses, as seen in debates over productivity measurement and inflation dynamics. Central Banks, with their unique technical expertise and infrastructural capacity, are uniquely positioned to mediate between economists and NSOs. Their role extends beyond data analysis: they must champion frameworks that respect both micro heterogeneity and macro measurement realities.

The path forward hinges on two priorities. First, economists must abandon the fiction that NA aggregates can—or should—be fully microfounded. Satellite accounts and top-down restrictions offer viable middle grounds, enabling theory-informed adjustments without dismantling NA coherence. Second, institutions like Central Banks should institutionalize cross-disciplinary dialogue, embedding NA statisticians in policy design to preempt interpretive pitfalls.

Ultimately, aggregation is not a peripheral technicality but the linchpin of macroeconomic storytelling. By taking measurement seriously—as a collaborative endeavor rather than a battleground of ideals—we can craft narratives that are both theoretically rigorous and empirically actionable. The alternative—a perpetuation of parallel discourses—serves neither science nor policy.

References

Balk, B. M. (2005). Divisia price and quantity indices: 80 years after. *Statistica Neerlandica*, *59*(2), 119-158.

Balk, B. M. (2016). The dynamics of productivity change: A review of the bottom-up approach (pp. 15-49). Springer International Publishing.

Balk, B. M. (2018). Aggregate productivity and productivity of the aggregate: Connecting the bottom-up and top-down approaches. In Productivity and Inequality IX (pp. 119-141). Springer International Publishing.

Balk, B. M. (2021). Productivity. Springer International Publishing.

Balk, B. M. (2024). Why is the Cobb-Douglas production function so popular?. *Evolutionary and Institutional Economics Review*, *21*(1), 1-20.

Blatnik, N., Morais, L. T.& Sola, P. (2022). New dataset on Distributional Wealth Accounts – Presentation of methodology and main results, DG-S Seminar, ECB, 2022.

Basu, S. and J. G. Fernald, 2002, Aggregate Productivity and Aggregate Technology, European Economic Review 46, 963-991.

Batty, M., Bricker, J., Briggs, J., Friedman, S., Nemschoff, D., Nielsen, E., ... & Volz, A. H. (2020). The Distributional Financial Accounts of the United States. Measuring Distribution and Mobility of Income and Wealth, 641-677.

Batty, M., Deeken, E., & Volz, A. H. (2021). Wealth inequality and COVID-19: Evidence from the distributional financial accounts.

Bernanke, B. (2024). Forecasting for monetary policy making and communication at the Bank of England: a review. *Bank of England Independent Evaluation Office*.

Bos, F. (2009). The National Accounts as a Tool for Analysis and Policy: History, Economic Theory and Compilation Issues. Economic Theory and Compilation, (2009).

Chrystal, K. A., & MacDonald, R. (1994). Empirical evidence on the recent behavior and usefulness of simple-sum and weighted measures of the money stock. *Review-Federal Reserve Bank of Saint Louis*, *76*, 73-73.

David, P. A. (1966). Measuring real net output: A proposed index. *The Review of Economics and Statistics*, 419-425.

Demski, J. S. (1973). The general impossibility of normative accounting standards. *The Accounting Review*, *48*(4), 718-723.

Dyson, F. J. (2007). Why is Maxwell's theory so hard to understand?. In *2nd European Conference on Antennas and Propagation (EuCAP 2007)*. Stevenage UK: IET.

Fisher F. M. (1969): 'The existence of aggregate production functions', *Econometrica*, 37 (4), pp. 553–77.

Fisher, F. M. (1971). Aggregate production functions and the explanation of wages: a simulation experiment. *The Review of Economics and Statistics*, 305-325.

Gorman, W. 1953. Community Preference Fields. *Econometrica* **21**:63–80.

Ijiri, Y. (1971). Fundamental queries in aggregation theory. *Journal of the American Statistical Association*, *66*(336), 766-782.

Hansen, L. P., & Heckman, J. J. (1996). The empirical foundations of calibration. *Journal of economic perspectives*, *10*(1), 87-104.

Hoover, K. D. (1995). Is macroeconomics for real?. *The Monist*, *78*(3), 235-257.

Janssen, M. (2005). *Microfoundations: A critical inquiry*. Routledge.

Kirman, A. (1989). The intrinsic limits of modern economic theory: the emperor has no clothes. *The Economic Journal*, *99*(395), 126-139.

Klein, L. R. (1946a). Macroeconomics and the theory of rational behavior. *Econometrica, Journal of the Econometric Society*, 93-108.

Klein, L. R. (1946b). Remarks on the Theory of Aggregation. *Econometrica: Journal of the Econometric Society*, 303-312.

Koopmans, T. C. (1947). Measurement without theory. *The Review of Economics and Statistics*, *29*(3), 161-172.

de Loecker, J. and J. Konings, 2006, Job Reallocation and Productivity Growth in a Post-Socialist Economy: Evidence from Slovenian Manufacturing", European Journal of Political Economy 22, 388-408.

May, K. (1946). The aggregation problem for a one-industry model. *Econometrica: Journal of the Econometric Society*, 285-298.

Meade, D. S. (2010). Why real value added is not my favorite concept. *Studies on Russian Economic Development*, *21*(3), 249.

Mirowski, P. (1989), 'The Measurement Without Theory Controversy: Defeating Rival Research Programs by Accusing Them of Naive Empiricism', Economies et Societes, 11, 65–87.

Nataf, A. (1948). Sur la possibilité de construction de certains macromodèles. *Econometrica: Journal of the Econometric Society*, 232-244.

Orchard, J., Ramey, V. A., & Wieland, J. (2024). *Using macro counterfactuals to assess plausibility: an illustration using the 2001 rebate MPCs* (No. w31808). National Bureau of Economic Research.

Pu, S. S. (1946). A note on macroeconomics. *Econometrica: Journal of the Econometric Society*, 299-302.

Ramey, V. A, 2019. "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives* 33(2): 89–114.

Reich, U. P. (2016). Value``added": Why consistency in aggregation is essential for global accounting standards, and how to achieve it. *Journal of Economic and Social Measurement*, *41*(3), 307-327.

de Soyres, F., Gaillard, A., & Young, H. L. (2023). What is Measured in National Accounts?. *International Finance Discussion Paper*, (1375).

Sato, K. (1976). The meaning and measurement of the real value added index. *The Review of economics and Statistics*, 434-442.

Sahm, C. R., Shapiro, M. D., & Slemrod, J. (2012). Check in the mail or more in the paycheck: does the effectiveness of fiscal stimulus depend on how it is delivered?. *American Economic Journal: Economic Policy*, *4*(3), 216-250.

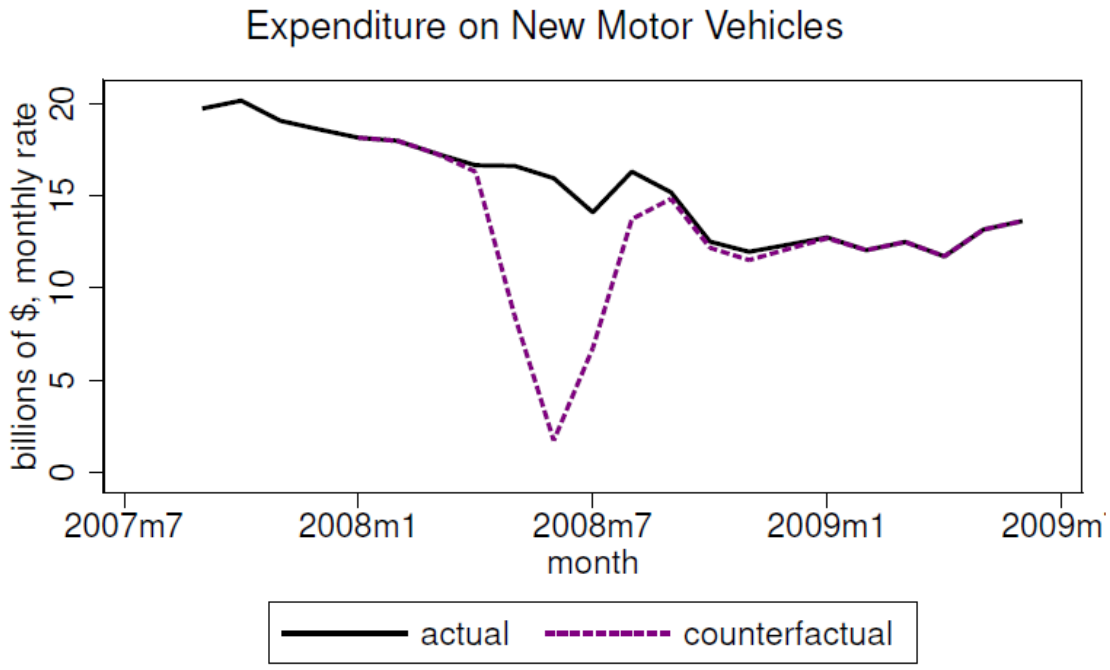
Schlicht, E. (2012). *Isolation and aggregation in economics*. Springer Science & Business Media.

Shaikh, A. (1974). Laws of production and laws of algebra: the humbug production function. *The review of economics and statistics*, 115-120.

Shoven, J. B., & Whalley, J. (1992). *Applying general equilibrium*. Cambridge university press.

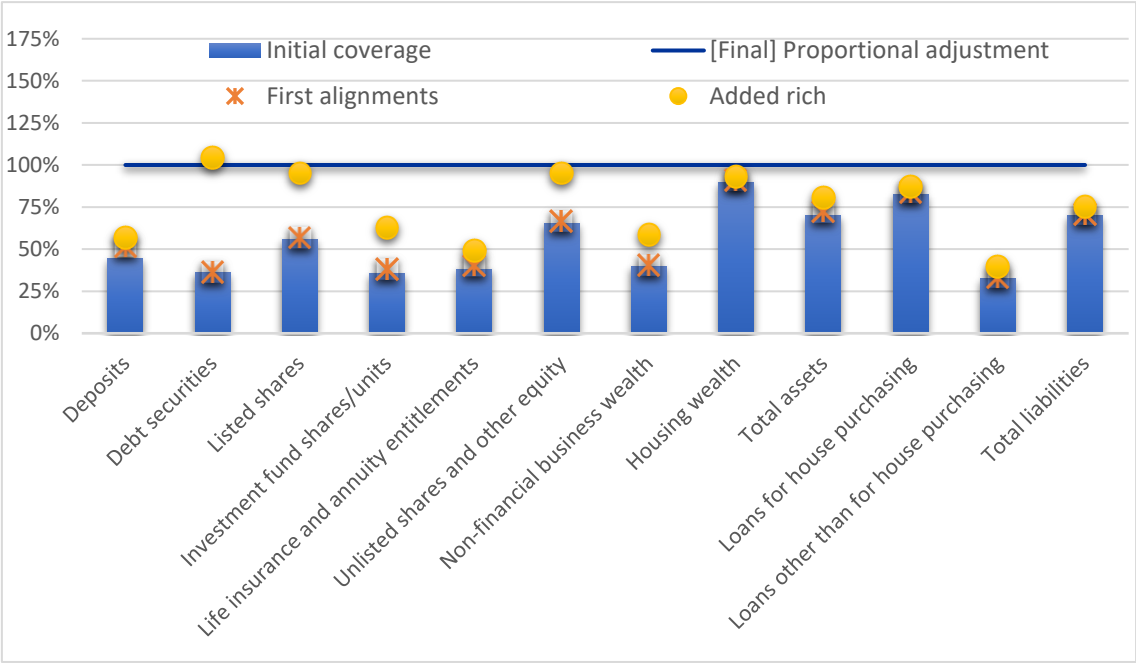
van Daal, J., & Merkies, A. H. Q. M. (1988). The problem of aggregation of individual economic relations; consistency and representativity in a historical perspective. In *Measurement in Economics: Theory and Applications of Economics Indices* (pp. 607-637). Heidelberg: Physica-Verlag HD.

Figure 1: Expenditures on New Motor Vehicles: Actual vs. Counterfactual (Orchard et al., 2024)



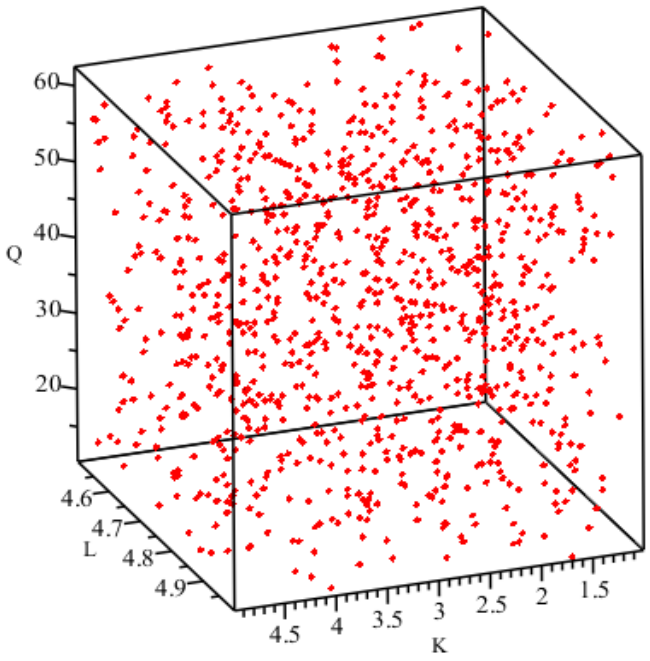
Source: Orchard et al. (2024, Figure 1)

Figure 2: Coverage ratio improvements for EA aggregate before proportional adjustment in DWA



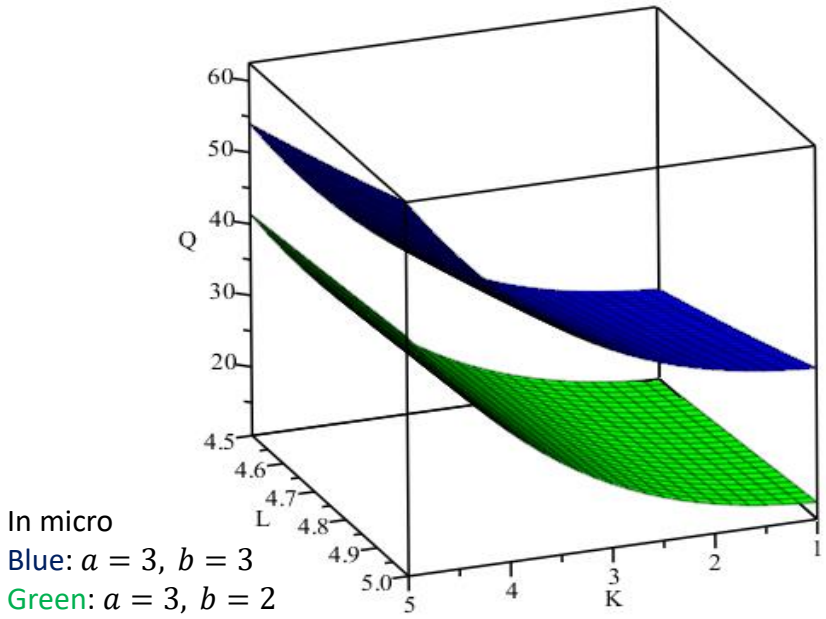
Source: Blatnik et al. (2022)

Figure 3 Closed and open aggregation under profit maximization



Note: both firms maximize profit under the production function $q_i = k_i^a + l_i^b,\ i = 1,\ 2$ with no restrictions on a and b. The aggregates Q, K and L are the simple sum of their micro counterparts.

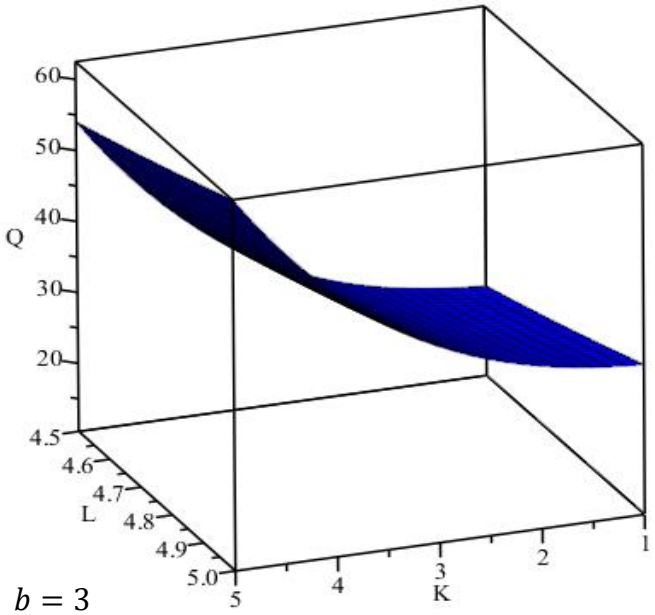
Figure 4: Closed and open aggregation under profit maximization and parameter restriction



In micro
Blue: $a = 3,\ b = 3$
Green: $a = 3,\ b = 2$

Note: both firms maximize profit under the production function $q_i = k_i^a + l_i^b,\ i = 1,\ 2$ with $a = 3, b = (2,3)$. The aggregates Q, K and L are the simple sum of their micro counterparts.

Figure 5: closed aggregation under profit maximization, parameter restriction and CRS

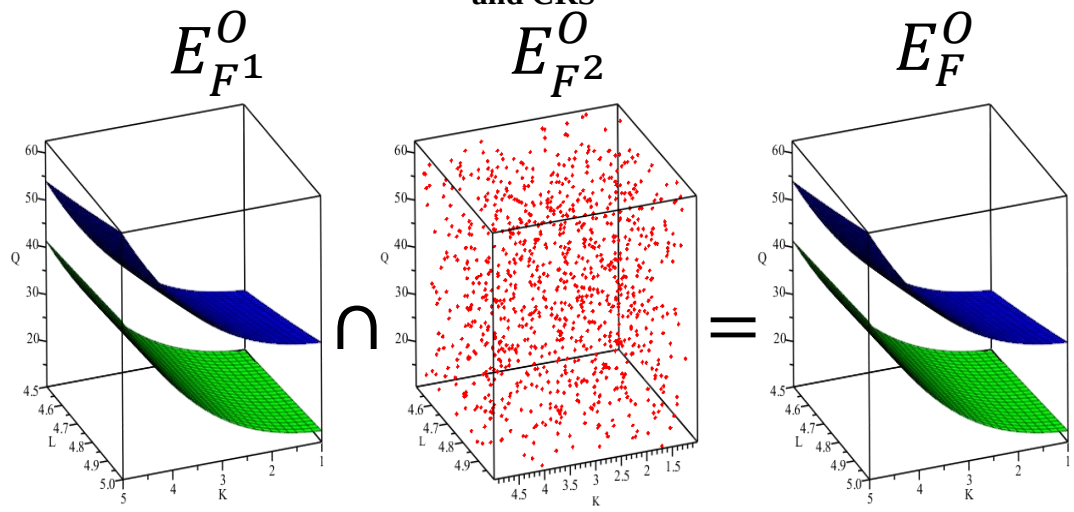


In micro

Blue: $a = 3,\ b = 3$

Note: both firms maximize profit under the production function $q_i = k_i^a + l_i^b,\ i = 1,\ 2$ with $a = 3, b = (2,3)$ and CRS. The aggregates Q, K and L are the simple sum of their micro counterparts.

Figure 6: open aggregation under profit maximization, parameter restriction and CRS



In micro

Blue: $a = 3,\ b = 3$

Green: $a = 3.\ b = 2$

Note: both firms maximize profit under the production function $q_i = k_i^a + l_i^b,\ i = 1,\ 2$ with $a = 3, b = (2,3)$ and CRS. The aggregates Q, K and L are the simple sum of their micro counterparts. National

accounts first aggregate partial equilibria separately to the macro level, denoted as $E_{F^1}^{O} = \alpha(E_{f^1})$ and $E_{F^2}^{O} = \alpha(E_{f^2})$. The open aggregation is formed as $E_F^O \coloneqq E_{F^1}^{O} \cap E_{F^2}^{O}$.

Appendix A The definition of closed and open aggregation

Micro equilibrium (eg individual consumption functions restricted by micro theory - $m = 2N$)

$$f(x) = 0,\ x \in R^m$$

Macro equilibrium (eg aggregate demand restricted by macro theory - $n = 2$) as a dimension reduction

$$F(X) = 0,\ X \in R^n,\ n < m$$

Mapping btw micro and macro (eg simple sum)

$$X = \alpha(x)$$

The set of micro equilibria

$$E_f \coloneqq \{x | f(x) = 0\}$$

The set of macro equilibria

$$E_F \coloneqq \{X | F(X) = 0\}$$

The aggregation problem is solved if the macro model and the aggregating function are selected such that

$$E_F = \alpha(E_f)$$

This is only a bottom line to build the valid microfoundations (Hoover, 2023).

Partitioned world: (partial) micro eq (eg f^1 covers the individual income and f^2 covers the individual consumption)

$$f^1(x) = 0, f^1 \colon R^m \twoheadrightarrow R^k,\ 1 \le k < m$$

$$f^2(x) = 0, f^2 \colon R^m \twoheadrightarrow R^{m-k}$$

(Partial) macro eq (eg F^1 covers the aggregate income and F^2 covers the aggregate consumption)

$$F^1(X) = 0, F^1 \colon R^n \twoheadrightarrow R^h,\ 1 \le h < k$$

$$F^2(X) = 0, F^2 \colon R^n \twoheadrightarrow R^{n-h}$$

F^1 and F^2 are taken to *subsume* f^1 and f^2 respectively

The corresponding sets of micro equilibrium

$$E_{f^1} \coloneqq \{x | f^1(x) = 0\}\ and\ E_{f^2} \coloneqq \{x | f^2(x) = 0\}$$

Closed aggregation: *restrict* the micro world then *aggregate* such that

$$E_F = \alpha(E_{f^1} \bigcap E_{f^2})$$

Open aggregation: *aggregate then restrict* the macro world (Schlicht, 1985)

$$E^o_{F^1} = \alpha(E_{f^1})$$

$$E^o_{F^2} = \alpha(E_{f^2})$$

$$E^o_F \coloneqq E^o_{F^1} \bigcap E^o_{F^2}$$

Appendix B Economic vs. axiomatic approach in the index number – epitome of two aggregations

- Index number problem: separate the price and quantity concepts in the macro

  Top-down (Milana, 2010; Reich, 2001) – unobservability of p and q in the micro level → infer the macro P (index) via identifying Q (index, eg real GDP) → circularity problem

  Bottom-up (Balk, 2009) – aggregating quantities of different commodities is impossible → aggregating indirectly by prices/values → dimensionality problem in interpretating P and Q indices

- Axiomatic approach – Euclidean geometry – open aggregation

  Impose aggregative axioms of the natural-intuitive princples on the macro P and Q (Hillinger, 2008; Donnell, 2012; Balk, 1995):
  Eg, P has a property of the inflation metric: if all prices change in the same proportion, the index must change in that proportion too (linear homogeneity axiom)

- Economic approach – rationalization – closed aggregation

  Characterise the dual changes of micro p or q when discussing q or p changes by micro theory (Milana, 2010; Balk, 1998):
  Eg, input price index is the ratio of cost functions when fixing output

- Deficiency of econ approach in capturing the micro reality – eg welfare (Hillinger, 2008; Jorgensen, 1990; Slesnick, 1998):

  Homethetic utility

  Aggregated indices instead of micro data used

  No generally accepted social welfare function

  Otherwise complexity deterring statistical agencies

- Inconsistency btw restrictions in approaches (Rudd, 2024):

  Product test in axiomatic approach: nominal consumption (change) divided by the price index is equal to the quantity index, and vice versa ← approximated by economic approach only under representative agent (Diewert, 1976)

- Do those natural-intuitive principles imposed by the Axiomatic Approach of index number theory objectively characterize the macro (data) relationship?

- Index number problem reflects the suffering of the circularity problem - 'problem of nomic measurement' (Chang, 2001):

  (1) We want to measure quantity *X*.

  (2) Quantity *X* is not directly observable by unaided human perception so we infer it from another quantity *Y*, which is directly observable.

  (3) For this inference we need a law that expresses *X* as a function of *Y*, $X = f(Y)$.

  (4) The form of this function *f* cannot be discovered or tested empirically, because that would involve knowing the values of both *Y* and *X*, and *X* is the unknown variable that we are trying to measure.

- Solution: consistency standard - not eliminates circularity but tightens the circle substantially. (Boumans, 2001)

- Same rationale behind filters, averaging, interpolation, seasonal adjustment…

- Circularity problem in the thermometer (Baumans, 2010)

  We require some material to expand linearly; yet we cannot prove the linearity of our target material without assuming the linearity of some other material. We cannot judge the inaccuracy induced by the expansion of the containing material without a standard of measurement; but any practical standard is itself subject to inaccuracy induced by its containing material.

- Henri Victor Regnault discovered gas thermometers were more comparable than mercury ones:

  One of the greatest advantages which a gas offers is its large expansion, and the resultantly great sensitivity of the thermometer. Also, because of this great expansion, the disturbing influence of the varying material of the vessels passes into the background. … gas expands 146 times as much as glass. The expansion has therefore only a small influence upon the apparent expansion of the gas, and

its change with different sorts [of glass] is of negligible influence. … The choice of material for the vessel, that is the individuality of the thermometer, can only disturb this relation insignificantly: thermometers become comparable to a high degree. (Mach [1896] 1966: 188)

- Convergence: case study of $E_{f^1} = E_{F^1}^{O}$ – Divisia index:
- Axiomatic approach perspective:

  Theoretically line integral indices (eg Divisia) can tackle the index number problem.

  Rationale: temporal identity (Reich, 2001; Balk, 2005) - weights to be up to date with the measured price/quality variables (eg removing substitution and/or new product bias; chained index)

- Econ approach perspective:

  Rationalization of the assumption behind: prices and quantiteis are functions of continuous time ← deriving Divisa from cost-of-living (Konüs, 1924) and cost minimization of firm (Balk, 2005; Hillinger, 2007, 2008)

  Micro restrictions = Macro restrictions in the context of aggregation ← compatible explanatory goals and micro restrictions involved are minimalist

Appendix C Consistency-in-aggregation (additivity)

- Def: the order in which steps of aggregation are taken from some lower to a higher level is irrelevant for the result.
- In a nutshell, additivity is ensured only when relative prices are constant (Reich, 2015)
- Chaining conflicts with addivity: relative prices do change after a while (Reich, 2008)
- Resulting discrepancies: after fixing the non-additivity caused by chaining for DK, an change of average growth rate of real GDP by either 3.6 pp per year or of 4.4 pp over 35 years btw 1970 and 2005 (Reich, 2015)