What are we measuring?

Measurement and aggregation issues in economics, with an application to climate risks

Executive Summary

Authors: Eddie Gerba and Gireesh Shrimali

November 2025

Abstract

This paper reviews the twin challenges of measurement and aggregation in economics and the natural sciences, with climate risk as a guiding example. It synthesises a broad range of theoretical and empirical perspectives, tracing ideas from early systems theory to modern macroeconomic debates, and compares the approaches of economics, complexity science, and climate science to the micro–macro aggregation problem. Several key conceptual tensions are highlighted—most notably the "micro–macro gap"—and the limitations of traditional models when confronted with heterogeneity, deep uncertainty, and non-linear feedbacks are demonstrated, especially in the climate-risk context. It also reviews emerging methodologies and proposes integrated frameworks to combine micro-level detail with macro-level consistency. Finally, the paper outlines a roadmap for future research and policy, advocating interdisciplinary collaboration, improved data infrastructure, and adaptive modelling strategies to better capture climate change.

***Keywords**:* Micro-macro gap, open vs closed aggregation, microfoundations, climate risks
***JEL codes**: B41, C18, C80, E10*

Full Paper: Gerba, Eddie. and Shrimali, Gireesh, (2025) "What are we measuring?" Measurement and aggregation issues in economics, with an application to climate risks. Available in SSRN

Correspondence: eddie.gerba@bankofengland.co.uk

Acknowledgement

Eddie Gerba: Bank of England, London School of Economics, and University of Oxford-SFG. **Gireesh Shrimali**: Smith School of Enterprise and Environment, and Sustainable Finance Group at Oxford University.

The work has greatly benefitted from comments and suggestions by Max Huppertz, Junyi Zhao, Lukasz Krebel, Marcin Borsuk, and Oxford-CGFI Fellows. This project was inspired by the thoughtful discussions around the 2023 Hybrid Workshop on Microfoundations in Measurement and Theory.

Executive summary

Introduction

Measurement and aggregation are interlinked challenges at the heart of understanding complex systems in both economics and the natural sciences (Sonnenschein, 1972; 1973; 1982; Simon, 1962). At the most fundamental level, the problem can be framed as: ***How can myriad micro-level elements or actors be meaningfully combined into coherent macro-level quantities or dynamics, without losing essential information?*** (Simon, 1962). This question surfaces in economics as the classic *aggregation problem* – how to derive reliable macroeconomic relationships from individual behaviour – and in fields like ecology or climate science as the problem of *coarse-graining* complex systems.

Our analytical review explores systematically these issues, from an interdisciplinary as well as inter-methodological angle. That is atypical in the literature and allows us to link theoretical (or conceptual) contributions across disciplines to empirical challenges and practical problems in climate prudential policy. To illustrate, we conceptually contrast closed to open aggregation and examine their implications for climate stress testing. We also discuss the inherent challenges of complexity and uncertainty in climate risk measurement, highlighting important trade-offs in any metrics or composite indicators, and provide a few (conceptually grounded) tentative solutions (e.g. scenario analyses, climate VaR, impact chains, and hierarchical models). We end the paper with some early suggestions for integrated frameworks and show how the proposed tools can be applied to specific policy considerations. We hope to substantially expand on this in subsequent papers.

We use climate risk as a recurring case study, while noting climate-specific nuances along the way. Climate risk – encompassing physical risks from climate impacts and transition risks from the shift to a low-carbon economy – is a domain where measurement and aggregation challenges are notably pronounced. Climate risk involves multi-dimensional, deeply uncertain, long-term processes that strain conventional statistical tools, and it requires combining insights from physics, economics, and other fields. By examining climate risk, we illustrate how general principles play out in practice, and how advances in one field (e.g. complexity theory) might inform another (e.g. macroeconomic stress testing for climate).

Key findings

First, aggregation issues are prevalent in economics, finance, as well as climate science. Aggregation is as much empirical as theoretical – deeply context-dependent. Every discipline has measurement protocols. For climate risk, think of how composite risk indices are built in vulnerability assessments (Fritzsche et al., 2014 – the “Impact Chain” approach used by GIZ); these effectively aggregate underlying factors with certain weights and formulas. These choices can introduce biases or hide variability. For example, a global climate risk index might combine economic losses, human fatalities, and ecological damage into one number per country, but that involves (explicitly or implicitly) value judgments about trade-offs between money, lives, and environment (Fleurbaey, 2009; Winsberg, 2012). Recently, **distributional national accounts** are being applied to reconcile micro data with macro totals. For instance, the US and EU now produce Distributional National Accounts that allocate aggregate GDP or wealth to population percentiles, ensuring the micro distribution sums to the official macro totals (Federal Reserve's Distributional Financial Accounts, ECB's Distributional Wealth Accounts). This requires adjusting micro data to match the aggregates. It's an example of modifying micro measurement to hit macro constraints.

Second, climate risk unfolds over very long horizons (decades to centuries) with deep uncertainty; therefore, forward-looking and multiple scenario approaches are crucial. For policymakers and planners, this is a communication challenge: how to summarise "climate risk" into a single indicator when it depends on human actions and deep future uncertainties? The answer is often: you can't and shouldn't. Instead, one uses stress test frameworks that acknowledge multiple possibilities. In the Bank of England's 2021 exploratory exercise, for example, banks had to report results under different scenarios (early policy action vs late action vs no action), and the regulator looked at the system's resilience under each. There wasn't one bottom-line number like in a capital stress test; rather, it was a range of outcomes and a qualitative assessment of vulnerabilities. This multi-scenario approach is essentially opening up the aggregation – not collapsing across scenarios but keeping them separate. It's an interesting case where, as mentioned earlier, providing a dashboard of indicators (one per scenario, plus perhaps a subjective judgment of plausibility) is more informative than any single composite metric.

Third, how we measure variables influences what relationships we observe at macro level. Aggregation problems can often be mitigated by better measurement – e.g., collecting more granular or comprehensive data (so we're not missing chunks that get imputed), or designing metrics that include distribution info (like reporting not just a single risk score but also concentration measures or tail stats). In climate risk measurement, this is evident: regulators ask not just for one aggregate like "climate VaR", but for a set of indicators – e.g. exposure metrics (like percentage of portfolio in certain risk categories) and stress test losses under scenarios. Together, these provide a mosaic of a bank's risk. If we only had one number, it would either obscure too much or have to be so conservative (to account for tails) that it wouldn't be useful for average conditions.

Fourth, given the difficulties outlined, researchers have developed various methods in different fields to improve how we aggregate information. The aim is to see what each discipline can learn from the others, and how, in tackling a problem like climate risk, a hybrid of these methods might be most effective. Table 1 provides a high-level comparison across a few dimensions (model type, treatment of heterogeneity, treatment of non-linearity/tails, data focus, conceptual tensions, emerging solutions) for three stylised approaches:

1. **General equilibrium** approaches (e.g. DSGE and standard metrics like CPI/GDP),
2. **Complexity Science** approaches (e.g. agent-based models and network models),
3. **Climate Science/Risk approaches** (e.g. IAMs and scenario analysis used in climate policy).

This table is not rigid – these fields overlap (economists are now also building ABMs; climate scientists use economic models, etc.) – but it highlights tendencies.

Table 1: Comparison of methodologies and conceptual approaches across disciplines.

Approaches - Dimension	General equilibrium (e.g. CGE, DSGE)	Complexity/Simulation (e.g. ABM, digital twins)	Climate Science Practice (e.g. scenario analysis)
Micro–Macro model	Representative agent or aggregate equations are common (assume a "typical" agent or use simplified macro relationships), sacrificing heterogeneity for tractability. (Most economic models until recently imposed aggregation methods differing from index-number practices used in data.)	Agent-based models and network simulations explicitly model many diverse agents and their interactions, letting macro properties emerge (no representative agent). There isn't a single closed-form "macro equation" – the model generates aggregate outcomes via simulation.	Integrated Assessment Models (IAMs) often use a top-down representative agent economy; however, impact models and risk assessments increasingly combine multidisciplinary modules (e.g. climate models + sector economic models) to capture differences across sectors/regions. Climate models themselves are aggregated at large spatial scales and then downscaled.
Treatment of Heterogeneity	Often assumed away or highly stylized (e.g. all consumers identical) to get closed-form results. Heterogeneity introduced only in special cases (two-agent models, etc.) – otherwise aggregates might behave erratically (per SMD theorem). Recent emerging work on HANK models is adding back some heterogeneity with numerical methods.	Fundamental to the approach: every agent can be different. The challenge of heterogeneity is tackled via computation rather than assumption. Emergent macro patterns (fat-tailed outcomes, cascades) arise naturally from diverse agent behavior. Complexity models embrace richness of types but may need reduction techniques (clustering agents) for interpretation.	Recognised as crucial: climate impacts are uneven, so analyses distinguish by region, sector, or population group. However, many policy models still used (until recently) a global or national average damage function. Newer climate risk frameworks (e.g. stress tests) segment data (by sector, geography) to keep heterogeneity visible. There is also heterogeneity in time: near-term vs long-term risks handled via scenario pathways.
Non-linearity & Tail Risks	Tended to linearise around equilibria for analytical convenience (e.g. linear approximations of models, assuming normal shocks). Extreme events often treated as exogenous	Embraces non-linearity: models include feedback loops (e.g. network cascades) and can generate power-law distributions of outcomes. Rare but massive events emerge in simulations. Rather than one outcome, an ABM	Non-linearity is explicit: damage functions are often non-linear (e.g. losses accelerate with temperature). Tipping points are studied, though hard to quantify. Scenario analysis captures some

	"shocks" rather than modelled. As a result, traditional aggregates can severely understate risk of rare disasters. (That said, some econ models do allow non-linear dynamics, but solving them analytically is difficult.)	yields a distribution of outcomes which can be examined for tail characteristics. Complexity theory explicitly studies critical thresholds, tipping points, and phase transitions – i.e. non-linear emergent phenomena.	non-linearity by considering qualitatively different futures. Moreover, use of extreme climate scenarios (like high-emissions RCP 8.5) brings tail-risk scenarios into planning. Still, some official estimates (like IAM-based social cost of carbon) arguably underweight tail risks.
Data & Measurement Focus	Relies on aggregate official data (GDP, CPI, etc.) which are top-down consistent but may mask micro variation. Micro data used separately (e.g. microeconometric studies) but often not integrated into macro models. There is a tradition of creating indices (CPI, etc.) – aggregating baskets into one number – reflecting value judgments (Fisher, 2005). Recently, more focus on using rich micro data to inform macro (e.g. central banks using big data on heterogeneity).	Utilises large micro-level datasets when available (e.g. detailed network data, firm-level data). Measurement is often granular: the state of every agent is tracked. To summarise results, relies on statistical analysis of simulation outputs (distributions, moments). Less reliant on official aggregate metrics, more on raw or synthetic data. However, complexity models sometimes face calibration issues – they produce "what ifs" more than precise fits to data.	Combines diverse measurements: physical metrics (temperature, sea level), economic metrics (losses, costs), and composite indices (vulnerability indices). The practice is to present multiple metrics instead of one (e.g. warming in °C, plus % GDP loss, plus specific risk indicators). However, for policy, composite indices (like climate risk rankings or a single "social cost of carbon") are often created, aggregating many factors into one score. Data gaps are acknowledged (e.g. missing asset-level data), leading to use of proxies and scenario data rather than purely historical data.
Conceptual Tensions	Micro vs macro: need to reconcile individual optimization with aggregate outcomes leads to paradoxes (fallacy of composition). Ontologically, often assumes a "representative" entity that may not exist. Has struggled with incommensurability of different theoretical constructs (national accounts vs micro concepts, as discussed). Also tension between theoretical elegance and empirical realism.	Reductionism vs holism: acknowledges that the whole can be more than sum of parts (emergence). Does not force one equilibrium paradigm – uses computational experiment to explore possibilities. But then faces interpretability issues: how to map complex simulation outcomes to simpler understanding or policy use? Also, results can be sensitive to agent rules chosen – raising questions of validation.	Different disciplines (climate science, economics, sociology) each have their own metrics and models – integrating them leads to incommensurability problems (e.g. economic cost vs human lives vs biodiversity loss). Often resolved by converting everything to monetary terms (for cost-benefit analysis), which is philosophically contentious. There's tension between short-term measurable risk vs long-term systemic risk (e.g. insurers focus on near-term, climate models on long-term), leading to an aggregation across

			time that discounts or neglects future risk.
Emerging Solutions	Developing heterogeneous-agent models with tractable summary statistics (e.g. using distribution's moments as state variables) to inform policy. Using *satellite accounts* to better align macro data with theory (e.g. separate accounting for natural capital or inequality). Increased use of micro data to validate macro models (e.g. granular data in central bank policy models). Essentially, economics is slowly moving toward embracing more complexity in models, aided by better computation.	Improving algorithms to coarse-grain models (e.g. find clusters of agents that can be treated as one without much error). Using machine learning as surrogate models to approximate ABM outcomes with simpler equations (to allow faster analysis or estimation). Integrating network metrics into policy frameworks (e.g. stress test triggers if network connectivity indicates vulnerability). Complexity science is also engaging with domain-specific data to calibrate ABMs more credibly.	IAMs are becoming more modular and stochastic, incorporating uncertainty explicitly (e.g. using Monte Carlo ensembles). Financial stress-testing frameworks are evolving to require granular data inputs from firms (so regulators can aggregate consistently). Proposals for hybrid modelling: e.g. run an ABM for one part of the economy (power sector) and link to a DSGE model for another part (the rest of economy), marrying detail with theory. Also, greater emphasis on common scenario sets (e.g. NGFS scenarios) so that different institutions' results can be compared apples-to-apples.

Fifth, understanding and improving measurement and aggregation isn't just an academic exercise – it has real consequences for policy and management in climate-related domains. In the paper, we discuss several areas where these issues play out in policy, and how better approaches can lead to better decisions, including: Financial regulation and systemic risk management, Macroeconomic policy and public investment, Corporate and portfolio strategy, Climate policy and integrated planning, Overarching issues of communication and trust, Managing policy trade-offs, and policy coordination on a global level. The climate risk challenge has accelerated improvements in these aspects. We can expect cross-fertilization – e.g., techniques from financial risk aggregation being applied to climate vulnerability assessment and vice versa.

Conclusion and Roadmap

We have seen how measuring and aggregating complex phenomena – such as economic welfare or climate risk – is fraught with challenges, yet crucial for sound decision-making. The way forward, underscored by recent advances, is to embrace complexity in our measurement and be nuanced in our aggregation. Key insights and takeaways include:

- ✓ **There is no single silver-bullet metric for climate or economic risk.** Instead, a portfolio of indicators is needed. Climate risk managers should consider physical risk, transition risk, tail scenario impacts, etc., separately before forming a composite view. Effective communication will involve conveying uncertainty ranges, not just point estimates.

- ✓ **Heterogeneity and distribution matter enormously.** Averages can mislead when distributions are broad or skewed. Future research should focus on developing better ways to incorporate distributional information into aggregate metrics – e.g. presenting inequality-adjusted aggregates or risk-adjusted aggregates (where a higher dispersion or tail risk inflates the effective aggregate risk measure). In climate risk terms, that might mean weighting metrics not just by mean outcomes but by concentration of risk (e.g. “40% of our exposure is accounted for by the top 10 polluting companies” is a distribution-aware statement).
- ✓ **Non-linear dynamics mean the sum of parts can behave in unexpected ways.** We must design models (and policies) that consider feedback loops across scales. The use of agent-based simulations and networks alongside aggregate models is a promising practice to test the robustness of aggregate predictions. For instance, if an IAM says “X% GDP loss”, but an ABM of firms shows potential collapse of network production beyond that, policymakers should account for that contingency (perhaps via scenario analysis or precautionary buffers).
- ✓ **Improving data quality and consistency is foundational.** Efforts like standardized climate disclosure (e.g. the new ISSB standards), open climate risk databases, and harmonized national accounting for climate impacts will greatly enhance our ability to aggregate meaningfully. Investment in data infrastructure (e.g. geospatial asset databases, climate-financial risk data hubs) will pay off by reducing the noise and bias in aggregate measures. Essentially, better micro data = more reliable macro aggregates.
- ✓ **Conceptual and normative clarity.** We should recognize what our aggregates represent and what they omit. GDP, for example, is not a welfare measure; adding natural capital accounting is one corrective. Similarly, a “1.5°C warming” target, while a useful aggregate goal, omits information about regional extremes – climate policy should incorporate complementary targets or bounds (perhaps something like “no region experiences >X°C increase” in addition to the global mean target). A future framework might set a vector of climate goals (temperature plus adaptation/resilience metrics) rather than a single number. In short, be clear about values and judgments embedded in aggregates.
- ✓ **Interdisciplinary collaboration.** Economists, climate scientists, and complexity theorists need to continue cross-pollinating methods. For instance, machine learning could be used to approximate the results of complex simulations in a formula that policymakers can use – essentially automating aggregation. Or insights from climate science about fat-tailed damage distributions could inform financial stress test scenarios to include more severe edge cases.

The road ahead for research and policy development includes:

- ➢ **Developing better theoretical aggregation theorems** for cases with near decomposability plus known exceptions (to guide when representative models are valid vs when ABMs are needed).
- ➢ **Building open-source simulation platforms** that allow users to plug in micro data and obtain aggregate risk distributions, lowering the barrier to sophisticated analysis.
- ➢ **Creating forums and standards** for sharing best practices on everything from how to aggregate climate scenarios across models to how to reflect model uncertainty in aggregated outputs (maybe by presenting ranges across models, as we discussed).

- **Encouraging policy exercises** like scenario gaming that explicitly address cross-sector aggregation – e.g., a national climate risk drill where different ministries (energy, agriculture, finance) input their sectoral assessments and a central team aggregates them to identify gaps (like something falling through the cracks at aggregate level).

In the end, tackling issues as sprawling as climate change or ensuring financial stability in a changing world is akin to solving a giant puzzle. Each piece (each dataset, each model, each sector) provides part of the picture. The job of researchers and policymakers is to fit these pieces together without forcing them into the wrong place or leaving gaps. That means sometimes aggregating, sometimes disaggregating, and always questioning whether the picture we see is true to the pieces that form it.

References:

Bank of England. (2021). *Climate Biennial Exploratory Scenario: Financial risks from climate change*. Bank of England.

European Central Bank. (n.d.). *Distributional Wealth Accounts*. https://www.ecb.europa.eu/pub/economic-research/research-networks/html/research_networks_dna.en.html

Federal Reserve Board. (n.d.). *Distributional Financial Accounts*. https://www.federalreserve.gov/releases/z1/dataviz/dfa/

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature, 47*(4), 1029–1075.

Fritzsche, K., Schneiderbauer, S., Bubeck, P., et al. (2014). *The Vulnerability Sourcebook: Concept and guidelines for standardised vulnerability assessments*. GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit).

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

Sonnenschein, H. (1972). Market excess demand functions. *Econometrica, 40*(3), 549–563.

Sonnenschein, H. (1973). Do Walras' identity and continuity characterize the class of community excess demand functions? *Journal of Economic Theory, 6*(4), 345–354.

Sonnenschein, H. (1982). Price adjustment and aggregate excess demand. *Econometrica, 50*(2), 539–547.

Winsberg, E. (2012). Values and uncertainties in the predictions of global climate models. *Philosophy of Science, 79*(5), 830–841.

The Smith School of Enterprise and the Environment (SSEE)

SSEE was established with a benefaction by the Smith family in 2008 to tackle major environmental challenges by bringing public and private enterprise together with the University of Oxford's world-leading teaching and research.

Research at the Smith School shapes business practices, government policy and strategies to achieve net zero emissions and sustainable development. We offer innovative evidence-based solutions to the environmental challenges facing humanity over the coming decades. We apply expertise in economics, finance, business, and law to tackle environmental and social challenges in six areas: water, climate, energy, biodiversity, food, and the circular economy.

SSEE has several significant external research partnerships and Business Fellows, bringing experts from industry, consulting firms, and related enterprises who seek to address major environmental challenges to the University of Oxford. We offer a variety of open enrolment and custom Executive Education programmes that cater to participants from all over the world. We also provide independent research and advice on environmental strategy, corporate governance, public policy, and long-term innovation.

For more information on SSEE please visit: www.smithschool.ox.ac.uk

Oxford Sustainable Finance Group

Oxford Sustainable Finance Group are a world-leading, multi-disciplinary centre for research and teaching in sustainable finance. We are uniquely placed by virtue of our scale, scope, networks, and leadership to understand the key challenges and opportunities in different contexts, and to work with partners to ambitiously shape the future of sustainable finance.

Aligning finance with sustainability to tackle global environmental and social challenges.

Both financial institutions and the broader financial system must manage the risks and capture the opportunities of the transition to global environmental sustainability. The University of Oxford has world leading researchers and research capabilities relevant to understanding these challenges and opportunities.

Established in 2012, the Oxford Sustainable Finance Group is the focal point for these activities.

The Group is multi-disciplinary and works globally across asset classes, finance professions, and with different parts of the financial system. We are the largest such centre globally and are working to be the world's best place for research and teaching on sustainable finance and investment. The Oxford Sustainable Finance Group is part of the Smith School of Enterprise and the Environment at the University of Oxford.

For more information please visit: sustainablefinance.ox.ac.uk/group