The Data Revolution, and Its Uses, in International Trade

By David Atkin and Benjamin Faber

In recent decades, economists studying international trade have gained access to an unprecedented volume and variety of data. These data have provided new insights and a more granular understanding of the mechanics of trade. Analyzing a large corpus of papers in international trade, we document that the nature of research has shifted alongside this explosion of information. In the early 2000s, empirical work typically documented new patterns, estimated the impacts of policies or tested model predictions. More recently, scholarship has shifted toward a more integrated approach, in which theory and empirics are tightly interwoven, particularly in the form of quantitative modeling. This development is surprising, in the sense that we might have expected the data revolution and the contemporaneous credibility revolution to have increased the share of primarily empirical papers—although the markers of these two revolutions are clearly evident in the rise of causal inference within empirical work in international trade. This article reviews these developments, assessing the strengths and limitations of different modes of inquiry, and plotting a path forward to best harness the growing richness of data within the field of international trade.

Over the past 25 years, the field of international trade has been transformed by an explosion in data availability and computing power. Researchers today have access to datasets as granular as transaction-level trade flows between establishments across and within borders, order-level price quotes, and high-resolution geospatial satellite imagery. Alongside this greater granularity, digitization has made it increasingly possible to link together administrative datasets to provide a more complete picture of the economy. As we document by analyzing a large corpus of publications, this data revolution has opened new lines of inquiry and enabled more rigorous testing of theoretical predictions. It has also, perhaps surprisingly, been accompanied by a dramatic rise in the use of quantitative modeling in international trade. These developments have enriched our understanding of many facets of the trade literature. However, they have also introduced new challenges. For example, the temptation to complicate models to make the most of new granular data sources at the cost of obscuring the key economic forces at play; or conversely limiting the scope of an empirical analysis because not all outcomes of interest can be captured within the same quantitative framework. More generally, the data revolution has naturally pushed the field towards topics where data is abundant and away from those where it is not. This article offers a critical survey of how the data revolution, and the changing ways the field uses data, have spurred and reshaped research in international trade, what we have learned, and future directions for the field in this light. To ensure the survey remains concise, we focus on research in international trade and, with a few exceptions, leave aside the economic geography and international macroeconomics literatures which have experienced their own data revolutions.[1]

We structure this article around an exploration of the types of analysis conducted in the trade field and how these patterns evolved alongside the data revolution. We start in Section I by proposing a taxonomy of research in international trade. At a broad level, we divide papers into

[1]The boundary between fields is often muddy, particularly regarding the trade and economic geography literature. One distinction is that much of the trade literature treats labor as immobile, as it mostly is across countries that trade models typically focus on. In contrast, economic geography typically focuses on the location of economic activity within countries and sees this mobility as a central force.

three categories: those where the primary contribution of the paper is theoretical, those where the primary contribution is empirical, and those that combine theoretical and empirical contributions. We further divide the second and third categories based on the nature of the empirical exercises in the paper. For primarily empirical papers, we distinguish between those that are descriptive in nature, those that estimate the impacts of policies or shocks, those that test predictions of (existing) trade models, and those that estimate key parameters and elasticities to better understand the implications of those models. For papers combining theoretical and empirical contributions, we distinguish between those where data are used as motivating evidence for a modeling exercise, those where data are used to test the predictions of a new model, those where the model is mainly used to guide and derive the empirical specifications, those where data are used to quantify (estimate or calibrate) the model for counterfactual analysis, and those where the model is used to derive a set of sufficient statistics that can be measured in the data. Throughout, we provide illustrative examples from the literature—focusing on topics where the data revolution has provided new insights or a deeper understanding of the mechanics of trade, as well as highlighting shortcomings.

With this taxonomy in hand, in Section II we utilize large language models (LLMs) to categorize a database of all trade-related papers published in leading economics journals between 2000 and 2025. The most striking finding is the rise of work that combines both theory and data driven by an increasing number of papers where data is used to quantify a model for counterfactual analysis—what we label the 'quantitative turn' in international trade. In terms of the share of work in international trade, the growth of this combined category has come at the expense of primarily theoretical work. Suggestive of the hypothesis that the field increasingly believes a theoretical contribution must be tightly linked to data to qualify as a major contribution, this pattern is even more stark in the top-5 journals compared to field journals. Over the study period, page lengths and coauthor counts have also increased markedly—from around 1.5 to 3.5—as might be expected as we move away from specialization on either the empirical or theoretical side.

In contrast to pure theory, the share of primarily empirical papers has remained steady between 2000 and 2025. We conjecture that absent this quantitative turn, it would have risen on the back of the data revolution and the concurrent credibility revolution that have shifted the gravity of other applied fields towards reduced-form empirical work. The influence of this credibility revolution is apparent within the body of primarily empirical papers from the pronounced rise in the share of papers estimating the impacts of policy change or shocks, as well as the increasing discussion of identification and use of techniques such as differences in differences and IV, again revealed via LLMs.

In Section III, we go beyond describing the evolution of the field and ask—in the context of the rise of causal inference and the quantitative turn that have come to dominate the trade literature—how can we use data and theory more effectively? Among other things, here we argue that researchers often have different goals that may not be best served by following a cookie-cutter approach to quantification—posing a policy question, writing down and estimating a model with relevant features, and then using counterfactuals to provide answers to two decimal places. But in cases where accurate counterfactuals to policy questions is the central contribution, we discuss how these answers can be made more credible through recent advances in model testing and the greater use of causally identified moments within quantitative frameworks. Finally, we conclude by prognosticating on the path of future research, given both the new types and sources of data that are becoming available, and the rapid advances in digitization and artificial intelligence (AI).

I. A taxonomy of empirical work in international trade

We first categorize the international trade literature into three broad categories based on whether the primary contribution or contributions of the paper are theoretical, are empirical, or are a combination of the two. We further divide the latter two categories into subcategories. To help clarify the classification, we discuss prominent examples within these subcategories. In contrast, as this review focuses on the data revolution in international trade, we do not devote space to discussing or further categorizing the primarily theoretical literature in this review. This taxonomy serves several purposes. First, we use it to illustrate the diversity of methodologies used in the trade literature. By way of examples, we highlight many areas where empirical research has been particularly fruitful at generating a deeper understanding of the field and inspiring new theoretical work—i.e. what has the data revolution in international trade taught us? Of course, these examples only provide a window into the many advances in the field brought about through the greater availability of data. Second, the taxonomy serves as a framework for analyzing the evolution of empirical work in the trade literature that we turn to in Section II. In particular, we use the taxonomy below as the instruction set for an LLM that reveals the evolution of published work in international trade. Finally, the taxonomy is helpful in crystallizing some of the strengths and potential shortcomings of different approaches that we discuss in Section III.

A. *Primarily Empirical Work*

The first strand of work in the trade literature is what we term primarily empirical work. This takes four distinct forms. The first subcategory seeks to document new patterns in the data or relationships and mechanisms. A seminal example comes from McCallum (1995) who documents the implicit size of the Canada-US border. Here the empirical economist often hopes that data leads theory, with the reduced-form findings potentially inspiring additional modeling work. The second subcategory of primarily empirical work involves estimating the causal effects of policies or shocks, for example Frankel and Romer (1999)'s work estimating the effect of trade on growth. The third subcategory involves testing the predictions of existing models, for example the classic literature testing the Heckscher-Ohlin trade model by scholars such as Leontief (1953), Bowen, Leamer and Sveikauskas (1987), and Trefler (1995). Closely related, a fourth subcategory once again speaks to existing models to estimate parameters or other empirical moments of interest, for example the large and growing literature estimating trade elasticities. These categories are not mutually exclusive—a detail that our classification exercise in Section II will reflect. In particular, both testing an existing model and estimating a key parameter in a model often go hand in hand. For example, Goldberg and Maggi (1999) both seek to test Grossman and Helpman (1994)'s "Protection for Sale" model as well as to estimate the model parameter that determines the weight governments put on lobbying contributions. But what distinguishes these approaches from pure theory, or papers that combine both theoretical and empirical contributions, is the lack of any claim to making a novel theoretical point. As has occurred across the applied fields in economics over the last 25 years, the increasing amounts of data and computation power as well as new empirical techniques originating in labor economics, development and industrial organization have greatly expanded the scope and richness of such work within international trade.

Empirical A: Documenting new patterns and associations

The first subcategory we discuss is primarily empirical work that aims to document patterns in the data, establish causal relationships or provide evidence for certain mechanisms without committing to a specific theoretical model.

Heterogeneous Firms and Trade: Perhaps the best example of where newly-available data has transformed the trade field is the voluminous literature documenting new facts about firm-level trade, rather than country- and sector-level trade. In the 1990s, researchers gained access to microdata on larger plants and firms in several countries that could be combined with data on trade participation or exposure. These data revealed a variety of patterns, including the existence of large producer heterogeneity even within narrow industries, the fact that only some firms export, and the fact that exporters are more productive than non-exporters (e.g., work by Bernard and

Jensen (1995) and Tybout and Westbrook (1995)). These properties of the data were not well explained by traditional trade theory, because of its common assumption of a representative firm within industries.

Particularly consequential were Roberts and Tybout (1997) and Bernard and Jensen (1999) who documented that more productive firms self-select into exporting, and Pavcnik (2002) who showed that trade liberalization induces market-share reallocations toward more productive firms. These empirical findings stimulated major theoretical advances in the field of international trade. Most notably, Melitz (2003) formalized these selection effects of trade in the presence of firm heterogeneity. Arguably, no paper over the last 25 years has had a larger influence on how economists think about the impacts of international trade. Bernard et al. (2003) also deserve mention for incorporating firm heterogeneity into Eaton and Kortum (2002)'s model structure that both rationalizes the main empirical regularities of trade flow data in a parsimonious manner and makes quantification feasible.

This literature further blossomed in the 2000s both because these theoretical contributions stimulated further empirical research and because such firm- or establishment-level data became available for more countries, and in some cases could be matched to transaction-level trade flows in customs microdata as well as other administrative datasets, such as social security rolls. These data proved a rich hunting ground for establishing additional facts about firm heterogeneity in international trade, both in terms of which destinations firms serve as in work by Eaton, Kortum and Kramarz (2011) and Dickstein and Morales (2018), which products firms sell as shown by Goldberg et al. (2010) and Bernard, Redding and Schott (2011), how firm heterogeneity is related to output quality as in studies by Verhoogen (2008) and Kugler and Verhoogen (2012), and which workers at which firms are impacted as explored by Helpman et al. (2017)—although many of these papers combine new empirical findings with theoretical modeling, a category of papers we discuss in Section I.B.

While trade flow data are very granular, serious measurement issues remain when trying to estimate objects such as productivity, markups, product variety and quality. Here, richer administrative data including prices and quantities of total output at the level of specific products as well as the corresponding inputs (e.g. De Loecker et al. (2016), Garcia-Marin and Voigtländer (2019) and de Roux et al. (2021)'s work) and tailored surveys (such as those fielded by Atkin, Khandelwal and Osman (2017)) have provided a deeper understanding, as have improved empirical methodologies such as the use of event studies, instrumental variable strategies and randomized control trials in these papers.

While the constructive interplay between data and theory in this space is highlighted above, there are also examples of theory not following the data. For example, Pavcnik (2002) provides evidence for trade liberalization affecting aggregate productivity through selection, but her paper finds an even bigger role for trade raising within-firm productivity than for across-firm selection effects—in line with other influential work on the effect of trade on within-firm productivity (e.g., Topalova and Khandelwal (2011)); yet that has garnered less theoretical interest in the subsequent years.[2]

Indirect exposure to trade: In recent years, administrative data that tracks transactions between domestic firms has become available for a number of countries, typically collected as part of the tax system. These data have led to an explosion of work documenting the true extent of exposure to trade occurring through domestic supply chains (e.g see Dhyne et al. (2021)'s work) as well as other features such as the skewness of trade networks and their stability (e.g., see the review by Bernard and Moxnes (2018)). Adao et al. (2020) combine these data with matched employer-employee and firm ownership records to calculate the nuanced exposure of individuals' income to trade, both through import competition and exports, as well as through the complementarities between different types of workers and imported capital.

[2]Notable exceptions include recent work by Mayer, Melitz and Ottaviano (2021) on selection effects within multi-product firms, or by Kotia (2025) on reducing x-inefficiency by professionalizing family firms in response to reductions in trade protection.

Measuring value-added in trade: With richer input-output and firm- or establishment-level datasets, researchers have moved beyond gross export totals to extract measures of domestic value-added. These measures reveal that gross trade overstates flows of true value-added because of multi-stage production chains that lead to the same value-added being counted anew every time it crosses a border (see Johnson and Noguera (2012) and Koopman, Wang and Wei (2014)'s papers) and provide visibility on the evolution of global value chains that is partially masked by gross trade flows.

Components of trade barriers: Microdata on the plumbing of trade (shipping and port records, customs times, firm-level mode choices) have made it possible to quantify non-tariff trade barriers. For example, time frictions that act like ad-valorem tariffs, especially for inputs and time-sensitive goods, e.g. work by Hummels and Schaur (2013) and Djankov, Freund and Pham (2010); information and network frictions which reduce search/matching costs, e.g., work by Rauch and Trindade (2002) on ethnic networks and Startz (2021) on buyer travel to source locations. These data suggest large impacts in tariff-equivalent terms of these non-standard components of trade costs that much of the literature previously abstracted from.

Measuring tariff evasion in customs data: The combination of customs records, clever empirical designs and survey evidence has illuminated sizable amounts of tariff evasion in developing countries, particularly where tariffs are high and products can be misclassified into nearby categories with lower rates (e.g. Fisman and Wei (2004)'s work exploiting inconsistencies in the importing and exporting country's trade records and Sequeira (2016)'s paper using surveys of customs facilitators). This evasion reduces the elasticity of trade to tariffs and is difficult to remedy when the return to corruption remains large, as shown by Chalendard et al. (2023) using administrative staffing records from inside a customs agency.

Empirical B: Estimating effects of policies or shocks

The second subcategory comprises empirical papers that are primarily focused on estimating the causal effect of a policy reform or economic shock on economic outcomes, such as GDP, trade flows, wages, income, employment, or product prices.

The Impacts of Trade Reforms on Prices: We note from the outset that the trade literature has produced only a small number of credible, and policy-relevant causal estimates of trade policy effects. It is perhaps most surprising that there is little credible work estimating the impacts of actual trade policy on prices, despite terms-of-trade considerations being central to understanding the welfare effects of trade policy—a point Goldberg and Pavcnik (2016) make forcefully in their handbook chapter. Recent exceptions include Amiti, Redding and Weinstein (2019) and Fajgelbaum et al. (2020)'s work which shows unexpectedly small responses of foreign prices to the Trump I tariff war (although recent work by Ganapati and Hottman (2026) suggests that because exporters offer quantity discounts, export prices did fall holding fixed quantities per shipment). For non-tariff barriers, evidence is even scarcer, with Atkin et al. (2024) using Argentinian transaction-level import license decisions an exception, documenting prices on average rising with quantitative restrictions. Similarly, our understanding of how trade policy changes are borne across firms at different points in supply chains and by consumers remains limited, though recent work is beginning to address this gap (see, for example, Gopinath and Neiman (2014), Flaaen et al. (2025*b*) and Flaaen et al. (2025*a*)'s contributions). We return to this shortcoming below when discussing trade elasticities and the profession's preparedness for the momentous 2025 trade policy shocks.

Local labor markets approach: A particularly influential strand of this literature has examined how trade shocks affect local labor markets. The availability of rich, comparable labor markets data for many regions within a country—such as India's National Sample Survey, Brazil's population censuses and matched employer-employee records, and the United States' Census Integrated Public Use Micro Samples—enabled researchers to exploit geographic variation in exposure to national-level trade shocks. Seminal contributions include those by Topalova (2007), who finds that Indian districts more exposed to tariff reductions experienced relatively slower poverty reduction; Kovak (2013), who shows that Brazilian regions facing larger liberalization-induced price declines saw correspondingly larger wage declines; and Autor, Dorn and Hanson (2013) who uncover that US commuting zones with production mixes more exposed to Chinese imports saw relatively higher levels of unemployment and relatively lower labor force participation and wage levels. Later work showed that these effects persisted and even grew many years after the shocks (see analysis by Dix-Carneiro and Kovak (2019) and Autor, Dorn and Hanson (2021)).[3] This local labor market approach has inspired extensions examining trade's effects on political outcomes, innovation, education, and health as well as a rich applied econometrics literature on shift-share research designs (reviewed by Borusyak, Hull and Jaravel (2025)).

For trade economists who have long hypothesized that trade creates winners and losers, the fact that import competition would lead to relative harms for factors used intensively in competing product is not surprising. But what was unexpected was that domestic factor mobility is sufficiently low that such impacts show up when comparing similarly skilled types of worker across different regions. These findings spurred a growing quantitative literature incorporating migration frictions into more standard models in order to generate such patterns (e.g. work by Caliendo, Dvorkin and Parro (2019)) as well as a valuable insight from Borusyak, Dix-Carneiro and Kovak (2023) that we may not expect substantial factor movements as the exposure to trade shocks is spatially correlated and migration strongly declines with distance.

Shock propagation and supply chains: The increasing availability of firm-to-firm transaction data has enabled researchers to trace how exogenous shocks propagate through production networks. Studies examining natural disasters—including those by Carvalho et al. (2021) and Kashiwagi, Todo and Matous (2021)—document how supply chain disruptions transmit across firms and borders, revealing the extent of production interdependencies that are masked by aggregate trade data. This work has deepened our understanding of the fragility and resilience of global value chains.

Spillovers from foreign direct investment: The literature on FDI spillovers to domestic firms has been substantially advanced by richer data. Early work by Javorcik (2004) and others examines whether multinational presence generates productivity spillovers to local suppliers based on input-output (IO) linkages or geography. The availability of firm-to-firm transaction level data allows us to see actual supplier relationships which are massively sparser than those inferred from IO linkages. Alfaro-Ureña, Manelici and Vasquez (2022) use these data to show that domestic firms becoming suppliers to multinationals experienced sustained improvements in employment, productivity, and sales to other buyers—impacts that would have been missed using prior approaches in their setting.

Empirical C: Testing model predictions

The third subcategory comprises empirical papers primarily focused on testing the predictions of existing theoretical models.

Testing theories of FDI: A particularly instructive example of how new data can challenge and refine theoretical predictions comes from the literature on foreign direct investment. Early work by Alfaro and Charlton (2009) used data on firm ownership and IO matrices to categorize vertical versus horizontal FDI, finding patterns that seemed consistent with theories emphasizing factor cost differences as drivers of vertical investment. When more granular data on zip code to zip code or internal multinational transactions became available, much of what would have been categorized as vertical FDI based on industry codes and IO tables did not in fact involve supply relationships

[3] Although, somewhat puzzlingly, as shown by Autor et al. (2025) using employer-employee panel data, the labor markets most affected by China in the 2000s actually had more employment than less exposed locations post 2010, albeit far less manufacturing employment. Relatedly, Borusyak, Hull and Jaravel (2022)'s revaluation of the China shock that explicitly uses an industry-shock level law of large numbers finds that the manufacturing employment effects are more robust than other local impacts.

(see work by Atalay, Hortaçsu and Syverson (2014) for the US, and Ramondo, Rappoport and Ruhl (2016) for international trade and US multinationals). This literature continues to evolve with Flaaen et al. (2025*b*) making use of establishment-by-product level output and input use data to allocate firm level imports and exports.

Testing the theory of comparative advantage: The data revolution has also enabled more rigorous tests of foundational trade theories. Bernhofen and Brown (2004) exploit Japan's transition out of autarky in the 1860s—and the availability of detailed historical price and trade data—to provide perhaps the cleanest empirical test of the theory of comparative advantage theory, confirming that Japan exported goods with low autarky prices and imported those with high autarky prices. Schott (2004) uses detailed product-level trade data to test predictions of both old and new trade theory regarding specialization patterns, while Nunn (2007) combines measures of contract intensity constructed from input-output tables with cross-country institutional data used heavily in the political economy and growth literatures to test whether contract enforcement shapes comparative advantage as predicted by incomplete contracting theories. More recently, Kikuchi (2025) shows that the relevance of skill abundance in shaping comparative advantage in skill-intensive sectors has vanished in the cross-country data, with the adoption of automation in skill abundant countries—inferred from robot usage data—providing an explanation.

Empirical D: Estimating elasticities and other key parameters

The fourth subcategory comprises empirical papers primarily focused on estimating a parameter of interest from existing theoretical models, including elasticities that describe how outcomes change as a function of another variable.

Trade elasticities, tariff pass-through and the 2025 policy moment: The recent resurgence of heavy-handed trade protection has exposed both the importance and the limitations of our empirical knowledge about key trade parameters. Two objects are particularly crucial for evaluating and informing these recent trade policy episodes: trade elasticities and tariff pass-through rates. Both of these were inputs used to calculate the Trump administration's "reciprocal tariff" needed to eradicate bilateral trade deficits. Debates around the formulae in 2025 revealed substantial disagreement among economists about the appropriate values—with economists outside the administration arguing that a pass through close to 1 was appropriate rather than 0.25 used by the administration, with the administration itself providing a range between 2 and 4 for the trade elasticity. There was limited discussion of the fact that in a widely-cited article, Fajgelbaum et al. (2020) directly report that the object of interest, the elasticity of trade value with respect to tariffs, equals 1.5 using actual tariff policy variation from Trump I tariffs, inconsistent with the commonly-assumed trade elasticities around 3.

This episode highlights a broader point—that despite the centrality of tariff elasticities to trade policy discussions, and trade elasticities to quantitative trade models and estimates of the gains from trade (as we will highlight in Section I.B), we still lack an empirical consensus on the size of these objects. In terms of heavily-cited estimates, Caliendo and Parro (2015) use cross-sectional tariff variation and ratios of trade flows to sweep out confounding price effects and find an average elasticity of -4.6. Boehm, Levchenko and Pandalai-Nayar (2023) use changes in most-favored-nation tariffs that are not targeted to a specific country to show that trade elasticities differ substantially across time horizons, with short-run elasticities around -0.8 converging to long-run values around -2.1 over 7–10 years. Teti (2024) documents erroneous interpolation issues in the commonly used World Integrated Trade Solution database produced by the UN, WTO, ITC and World Bank. When corrected for this data issue, the above estimates change to -6.2, 0.5, and 1.67 respectively. Further correction for sample selection bias changes them again to -4.48, -0.5 and -0.8, respectively. Shapiro (2016) provides another recent estimate, relying on changes in shipping costs within origin-destination pairs (shipping costs are reported in the customs records of some countries) and correcting for measurement error using an IV approach, and finds an overall trade elasticity of -8. While we should not expect the same elasticity estimate over different time horizons and different contexts, our policy advice and the performance of our parsimonious quantitative models will only be as good as our elasticity estimates. We are far from achieving consensus on these objects that are so central to the field.

This limited consensus has echoes of the discussion above about the surprising lack of evidence on how tariffs affect prices, despite terms-of-trade effects being central to evaluating trade policy. What we have almost exclusively comes from the very recent Trump I tariff episode. Speculatively, these shortcomings may arise from limited interest for empirical researchers to work on these somewhat narrow measurement topics and limited demand from more theoretically minded researchers for more credible inputs.

Industrial policy: 2025 also brought much discussion about the merits of industrial policy. A small recent literature has emerged documenting the impacts of infant industry protection and industrial policy on the growth of the supported sectors (e.g. work byJuhász (2018), Liu (2019), or Lane (2025)). However, the optimality of industrial policy hinges on estimates of external economies of scale and spillovers at the sector or firm level, about which we know relatively little, with the work of Bartelme et al. (2025) a notable exception.

Distance Elasticities: Disdier and Head (2008) provide the canonical meta-analysis of distance elasticities in gravity equations—i.e. how rapidly trade flows fall with log distance. They document that for many studies run over many years, the distance coefficient remains close to one. In fact, estimates from the 1870s-1960s are lower than in more recent periods despite a belief that the costs of distance have been falling with transportation, communication and logistics improvements.

B. Combining theory and empirics

The second strand of work in the trade literature comprises papers that combine theoretical and empirical contributions. We distinguish between five types of analysis that combine theory with data. In the first subcategory, data analysis is used to present motivating evidence or stylized facts that are central to informing the structure and assumptions of the theoretical model. In the second subcategory, the data analysis serves to test the predictions stemming from a new model that the paper develops, for example Bernard et al. (2012)'s work on modeling the effects of trade liberalization in the presence of multi-product firms. The third type of analysis uses theory to guide and derive the empirical specifications, as in Donaldson (2018)'s work on the consequences of railway expansion in India. In the fourth subcategory, the data are used to calibrate or estimate parameters needed to quantify a model and conduct counterfactual analysis, as for example in the pathbreaking work by Eaton and Kortum (2002). A final subcategory uses theory to identify a small set of moments in the data and/or key empirical parameters needed to evaluate the effects of a policy, without fully specifying the underlying structural model, as in Arkolakis, Costinot and Rodríguez-Clare (2012)'s result that for many trade models, the gains from trade depend only on the share of expenditure on domestic goods and a single trade elasticity. Again, these categories are not mutually exclusive. For example, Eaton and Kortum (2002) lay out stylized facts and Donaldson (2018) carries out a model validation exercise. Helpman, Melitz and Rubinstein (2008) and Bernard et al. (2003) provide other well-cited examples of papers spanning multiple categories. What sets these categories apart from the primarily empirical work we describe above is that these papers' main contribution does not clearly lie in the theory or the empirics alone but in the interplay between theoretical modeling and data analysis.

Combined A: Models with motivating evidence/stylized facts

The data analysis in this subcategory of papers is used to present motivating evidence, often in the shape of new stylized facts, that inform the structure and assumptions of the theoretical model.

Examples of this approach span many areas of the trade literature. Atkeson and Burstein (2008) document patterns in exchange rate pass-through across destinations that motivate their model of oligopolistic competition in international trade. Helpman, Melitz and Rubinstein (2008) use the prevalence of zero trade flows between country pairs to motivate incorporating firm heterogeneity and extensive margin decisions into gravity estimation. Fajgelbaum, Grossman and Helpman (2015) document that FDI flows are disproportionately between countries at similar income levels—a pattern consistent with a demand-driven "Linder hypothesis"—before developing a model that rationalizes this pattern. Indeed, many of the most influential quantitative papers in trade lead with stylized facts that both justify the modeling choices and provide targets for calibration.

Combined B: Models whose (new) predictions are tested

The data analysis in this subcategory of papers is used to test the predictions of the theoretical model.

Testing takes multiple forms in this literature. Prior to the data revolution and in its early years, the most common approach in the trade literature was performing comparative statics in the model and testing those sign predictions (where we construe comparative statics broadly to include cross-sectional predictions of the model). For example, Antras (2003) uses BEA data on the operations of US affiliates to test incomplete contracting theories of multinational firms, Hanson and Xiang (2004) test home market effects in a multi-industry model, Helpman, Melitz and Yeaple (2004) test predictions about which markets firms choose to serve via exports versus FDI, and Bernard, Redding and Schott (2011) examine how multi-product firms adjust their product mix in response to trade shocks.

Other papers conduct tests that are more structural in nature: Donaldson (2018) shows that a substantial share of the reduced from gains from Indian railroad expansion can be explained by the sufficient statistic outlined in the work by Arkolakis, Costinot and Rodríguez-Clare (2012) mentioned above. Adão, Costinot and Donaldson (2025) implement a testing procedure that draws on exclusion restrictions already assumed for identification and applies the test to Fajgelbaum et al. (2020)'s model evaluating the Trump I tariff episode. Redding and Sturm (2008) use an extension of Helpman (1998)'s model to test economic geography theories of market access.

A growing practice involves testing model fit using untargeted moments—data patterns that the model was not calibrated to match—with recent examples including work by Castro-Vincenzi et al. (2024), Dix-Carneiro et al. (2021), Dix-Carneiro et al. (2023) and Sotelo (2020). Thus far, few papers have followed a similar approach to test the validity of counterfactual policy analysis in quantitative models—an approach suggested by Kehoe (2005) with a more modern treatment developed in recent work by Adão, Costinot and Donaldson (2025). We return to this shortcoming in Section III below.

Combined C: Models used to guide specifications or identification

The model in this subcategory of papers is used to derive the specifications for the empirical analysis, such as regression specifications, or to discuss threats to identification.

Theory can guide specifications in several ways. The first three steps of Donaldson (2008)'s analysis derive estimable reduced-form relationships from an underlying model, providing clear interpretations for the estimated coefficients. Fajgelbaum et al. (2023) similarly use theory to derive specifications that reveal optimal fiscal policies in a spatial equilibrium. Redding and Sturm (2008) use Helpman (1998)'s model to derive empirical specifications linking wages to market access.

Theory can also guide identification by clarifying what variation is needed for causal inference. Costinot et al. (2019) work on home market effects in pharmaceuticals provides an example of a theoretical framework that motivates and allows interpretation of a quasi-experimental research design—absent the model it is not clear what is in the error term of the key estimating equation and so what exclusion restrictions an instrument must satisfy. Faber and Gaubert (2019) write down a model in which amenities specific to tourism travel provide exogenous shifters in local tourism development, guiding the choice of natural amenities for tourism as such a shifter (an instrumental variable) which is then combined with the structure of the model to estimate within- and cross-sector agglomeration economies.

Combined D: Models that are quantified

The data analysis in this subcategory of papers is used to estimate the key model parameters that govern the model's behavior—typically in order to perform counterfactuals or simply to quantify the importance of the forces captured by a model. In the trade literature, this is typically done through calibration where moments in the data are used to pin down key parameters through the lens of the model, or elasticities are taken from external sources. For example, structural gravity typically uses GDP data alongside exporter and importer fixed effects recovered from gravity regressions to determine the relative sizes and productivities of different countries and sectors. Many papers combine calibration with structural estimation where parameters are chosen to best fit the data through an explicit objective function such as maximum likelihood or GMM.

While computable general equilibrium models of trade have long been used by policymakers (see Shoven and Whalley (1984)'s overview), they fell out of favor in academic research. Their complexity meant it was often difficult to discern what features drove which outcome and they required a large number of parameters that were typically calibrated, with the resulting estimates not deemed credible. Eaton and Kortum (2002) breathed new life into quantitative work in trade by writing down a model that, drawing heavily on convenient features of the Fréchet distribution, generated rich heterogeneity in trade patterns from reasonable micro foundations while remaining parsimonious and tractable and, critically, amenable to estimation using observable data on bilateral trade flows, prices and geography. Dekle, Eaton and Kortum (2007) refined the usefulness of the methodology by showing that directly solving for proportional changes rather than levels of outcomes such as prices, wages, and trade flows further reduces the parameters requiring estimation as non time-varying objects such as technologies and preferences are swept out (here, closely echoing approaches in the earlier CGE literature of expressing equilibrium conditions in calibrated share form).

The Eaton Kortum agenda has become the default for pursuing many questions in trade, with highly-cited extensions including those by Simonovska and Waugh (2011) and Caliendo and Parro (2015). Related approaches have been used to estimate dynamic models of trade, such as work by Caliendo, Dvorkin and Parro (2019) and Galle, Rodríguez-Clare and Yi (2023), trade policy uncertainty as analyzed by Handley and Limão (2017), multinational production explored by Ramondo and Rodríguez-Clare (2013), or production networks examined by Eaton, Kortum and Kramarz (2011).

Once calibrated, these models serve two main purposes. First, they can quantify the relative importance of different mechanisms—for example, Hottman, Redding and Weinstein (2016) decompose the sources of firm size dispersion. Second, they enable counterfactual analysis of policy changes that have not yet occurred, as illustrated in Costinot and Rodríguez-Clare (2014)'s handbook chapter on trade and welfare.

Combined E: Models used to provide sufficient statistics

The approach in this subcategory of papers focuses on identifying a small set of key empirical parameters needed to evaluate the effects of a policy, without fully specifying the underlying structural model, and then uses data to estimate those parameters as credibly as possible.

The most influential recent example in the trade literature is the work of Arkolakis, Costinot and Rodríguez-Clare (2012) who show that the share of expenditure on domestic consumption and the trade elasticity is sufficient to calculate the gains from trade in a range of commonly-used trade models. This approach requires knowing only endogenous variables that are observable (the expenditure share on domestic consumption) and reduced-form elasticities, where these elasticities can be composites of structural parameters. In other words, the answer to the question what are the gains from trade is the same in several classes of structural models given these two sufficient statics—enhancing the robustness of the results both because of the isomorphism and because fewer assumptions are required to estimate these sufficient statistics than to estimate the full set of structural parameters from one specific model within the classes spanned. Additionally, the sufficient statistics often hold as a first-order approximation to a wider class of models, in which the reduced-form elasticities are endogenous objects rather than constant parameters. Examples where the authors both derive sufficient statistics before using data to estimate these objects include work by Adao, Arkolakis and Esposito (2024) on the spatial impacts of international trade shocks on neighboring localities, Kleinman, Liu and Redding (2024) on the links between economic dependence and political alignment, and Huo, Levchenko and Pandalai-Nayar (2025) on the contribution of global production network to international GDP comovement.

Examples of Influential Combinations of Theory and Empirics

We now discuss two examples of topics where the combination of theoretical and empirical work in trade has been highly successful at influencing the field.

Estimating the gains from trade: In what Head and Mayer (2013) refer to as "convergence", models featuring structural gravity became the workhorse approach to quantifying the gains from trade. This tradition followed seminal contributions including the work of Arkolakis, Costinot and Rodríguez-Clare (2012) discussed above as well as Chaney (2008), Helpman, Melitz and Rubinstein (2008), and Melitz and Ottaviano (2008)'s papers that extend the selection forces in Melitz (2003)'s model in various directions. This convergence provided researchers with a widely used class of core models and a clear "Cookbook" (Head and Mayer (2013)) for identifying the data moments and estimation procedures needed to quantify models for counterfactual analysis.

The data requirements for this literature centered on cross-country information on sectoral output, trade flows, and input-output structures—often obtained from projects such as the Global Trade Analysis Project (GTAP). The trade elasticity became the key parameter governing welfare calculations, spurring empirical work estimating this elasticity across sectors and contexts using varied approaches. We touch on the purely empirical aspect of this literature, and its lack of consensus, above. The quantitative framework has since been extended to incorporate imperfect competition (Arkolakis et al. (2019)), input trade, (Blaum, Lelarge and Peters (2018)), multinational production and FDI (Ramondo and Rodríguez-Clare (2013)), endogenous networks (Arkolakis, Huneeus and Miyauchi (2025)), and dynamic gains from trade (Boehm et al. (2024)).

Most of this quantitative literature has relied on model-based counterfactuals to quantify the gains from trade—using the structural gravity property that is common across the field's core models, enabling quantification using exact hat algebra (Dekle, Eaton and Kortum (2007)). Notable exceptions that quantify gains from openness outside this structure include Donaldson (2008)—who provides reduced-form estimates alongside model-based ones—and Atkin, Faber and Gonzalez-Navarro (2018) who use first-order expressions that are direct functions of reduced-form effects.

Costinot and Rodríguez-Clare (2014) provide an overview of this now-standard approach for estimating the gains from trade, exploring how large those gains are for many countries first using our simplest models, and then progressively complicating them. The goal here is not to produce a complete model of the economy that can produce indisputable numbers, but instead show how magnitudes change when new features are added, and allowing us to judge how complete the benchmark models may be by comparing magnitudes to more reduced form estimates from elsewhere or our priors. For example, the relatively small gains from going from autarky to the current levels of trade for China, even in the most complete model they explore (equal to approximately 4 years of China's GDP growth in the early 2000s), suggest a large role for dynamic gains from trade that are outside the class of models covered.

Distributional Implications of Trade: Beyond the prominent role that Melitz (2003) has played in the literature on the gains from trade detailed above, it has also fundamentally altered the way that trade economists think about and rationalize the distributional impacts of trade. Incorporating multiple types of labor into models of firm heterogeneity (as in the work of Costinot and Vogel (2010) and Sampson (2014)), considering unemployment and firm-specific wage premia due to efficiency wages (as in the study by Davis and Harrigan (2011)), or accommodating labor-market frictions and job search (as in papers by Helpman and Itskhoki (2010), Helpman, Itskhoki and Redding (2010)) have allowed the trade literature to explain observed changes in wage inequality within sectors and types of workers that were at odds with the Stolper-Samuleson forces that dominated earlier discussions of trade and inequality.

Understanding Multinational Location Decisions: A growing literature incorporates multinational production into quantitative trade models by combining detailed firm-level data with discrete-choice formulations of production location and market entry decisions. Data on bilateral trade flows, affiliate sales, plant locations, and firm-level organizational structures disciplines models in which firms choose portfolios of production locations and markets subject to trade costs, fixed costs, and organizational frictions. Building on Eaton and Kortum (2002)'s model, work by Ramondo and Rodríguez-Clare (2013), Ramondo, Rodríguez-Clare and Tintelnot (2015) and others shows how observed patterns of exporting and foreign affiliate activity can be mapped into choice probabilities over complex location configurations. More recent work has framed the question as a combinatorial discrete choice problem rather than utilizing probabilistic approaches. Antràs, Fort and Tintelnot (2017) leverage multinational firm-level data and a model of sourcing that restricts complementarities to operate globally to estimate a model in which firms select sets of production locations, with the theoretical structure making the empirical estimation tractable. Arkolakis, Eckert and Shi (2023) and Castro-Vincenzi et al. (2024) instead provide conditions under which their algorithms can find global optima quickly, allowing for locations to be complements or substitutes with each other in a less restricted manner—with the latter set of authors also incorporating the firm-level decision of what products to sell in which markets. Across these studies, the discrete-choice structure allows high-dimensional firm decisions to be tightly linked to observable data, enabling counterfactual exercises that quantify the roles of trade costs, fixed costs, and policy changes in shaping multinational production and global value chains.

Spatial equilibrium: Starting with the work of Allen and Arkolakis (2014) and Redding (2016), Eaton and Kortum (2002)'s modeling innovation has been extended to settings with factor mobility, treating regions within a country analogously to countries in trade applications, while allowing for labor mobility across regions. These modeling advances unlocked an explosion of quantitative analysis in economic geography and urban economics—the latter particularly through the seminal paper by Ahlfeldt et al. (2015)—that followed on the heels of the quantitative turn in trade.

The transfer of these tools to spatial economics has required different and more granular data than used to answer questions in international trade: data across multiple regions or municipalities within a country (or census tracts within cities) in addition to trade and migration flows among those regions (e.g., see the work of Fajgelbaum and Redding (2022)). Similarly, spatial questions often require additional elasticities. For example, estimates of migration elasticities across regions, and commuting elasticities in urban settings (as in work by Monte, Redding and Rossi-Hansberg (2018)), have played important roles in quantifying spatial equilibrium models.

One central feature that the application of the structural-gravity toolkit to spatial economics has highlighted is the presence of market failures in the decentralized equilibrium due to agglomeration and congestion externalities (e.g., see the work of Fajgelbaum and Gaubert (2020)). Thus, these models have allowed economists to better understand tradeoffs in regional policy, for example by quantifying the costs of internal barriers to migration allowing for both agglomeration economies and worker sorting across regions, as in papers by Bryan and Morten (2019), Tombe and Zhu (2019) and Lagakos, Mobarak and Waugh (2023).

II. The Evolution of Empirical Work in International Trade

To better understand how research in international trade has evolved alongside the data revolution, we analyze a comprehensive corpus of 1,421 papers in the field published in leading economics journals between 2000 and 2025. Our sample encompasses seven journals: the traditional "top 5" journals, the American Economic Review (193 papers), Econometrica (33 papers), the Journal of Political Economy (54 papers), the Quarterly Journal of Economics (90 papers), and the Review of Economic Studies (89 papers); the Review of Economics and Statistics (81 papers) that has been an important outlet for trade papers; and the leading field journal in international trade, the Journal of International Economics (881 papers). This selection allows us to examine whether methodological trends differ between generalist and specialist outlets. To ensure that the selected papers were in relevant fields, we restricted papers by JEL codes or keywords for journals that do not use these codes.[4] We manually pruned the resulting list to remove non-trade papers, predominantly finance papers (involving the trade in assets).

To classify each paper according to the taxonomy described in Section I, we employed two state-of-the-art large language models (LLMs): Gemini 3.0 High and Claude Opus 4.5. Each LLM read the first 2,500 words of each paper—which typically covers the abstract, the introduction, and early sections of the paper—and assigned it to one of the three main categories, along with the applicable subcategories for empirical and combined papers. To ensure reliability, we implemented a three-stage classification process. In the initial stage, each LLM provided a classification along with a brief three-sentence justification for its choice, quoting the authors regarding the paper's primary contributions where possible. In the validation stage, the LLM reviewed its own justifications for internal consistency and, where discrepancies were identified, re-read the paper and adjusted its classification as necessary. This process (of using two LLMs) yielded two sets of classifications, allowing us to assess inter-rater reliability between models. In the third step, an additional Opus 4.5 agent evaluated and arbitrated any cases where the two LLM classifications yielded different main category assignments. The online appendix contains the detailed prompts as well as the full corpus of papers and classifications.

The classification results from the two LLMs accord strongly. Before the first validation, the two models agreed on the primary category assignment for 88% of papers; after the first validation, this agreement rate rose to 92%. The arbitration agent then sided with validated Opus 4.5 output 66% of the time and the validated Gemini output 34% of the time.[5]This high degree of inter-model

[4]We restricted attention to codes F1: Trade, F2: International Factor Movements and International Business, F6: Economic Impacts of Globalization, and R1: General Regional Economics. For journals that did not use JEL codes, we mapped each JEL code to keywords using the AEA's JEL Classification Codes Guide and searched for these keywords in the title and/or abstract of the paper. Econometrica, the Journal of International Economics, and the Review of Economics and Statistics do not use JEL codes while several other journals did not use them until the 2010s, or used them in an inconsistent manner.

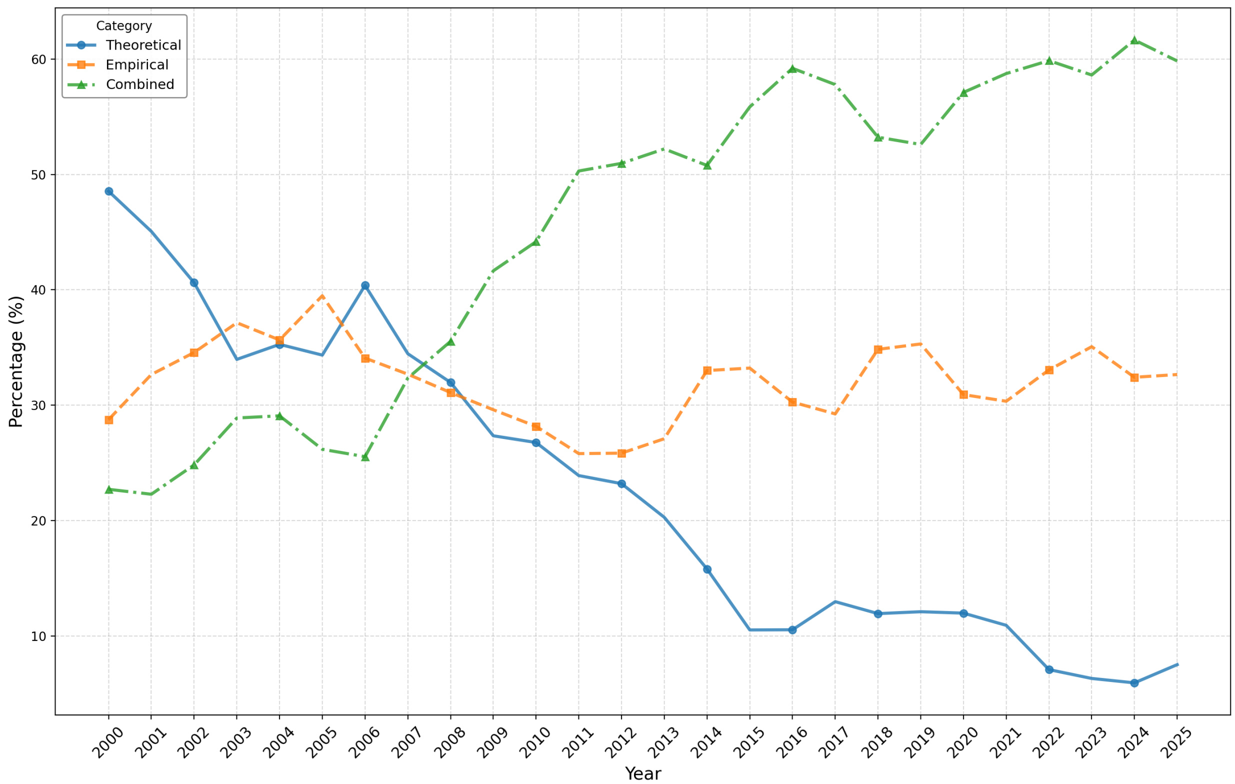
[5]The first validation process itself resulted in relatively few changes: Gemini adjusted its classification for 49 papers (3.4%), while Opus adjusted only 15 papers (1.1%). The final category distributions were virtually identical across models. Gemini classified 21.5% of papers as primarily theoretical, 31.7% as primarily empirical, and 46.7% as combining theory and empirics; Opus' corresponding figures were 20.8%, 32.3%, and 46.9%.

agreement provides us with some confidence in the reliability of our classifications. We follow the same three-step classification process described above to allocate papers with primarily empirical or combined contributions to the the sets of subcategories from Section I. As a single paper can use several approaches, we allow for papers to be classified into multiple subcategories (within the same broad category).

Our classification exercise reveals three major patterns related to the evolving usage of data and empirical evidence in the trade literature over the past quarter century.

A. The Decline of Specialization, and the Quantitative Turn in International Trade

Figure 1. : The Evolution of the Trade Field: Shares of Paper by Methodology, 2000–2025



Notes: Figure 1 plots the share of papers in the field of international trade published in seven leading journals between 2000 and 2025 by broad methodological category. Categorization done via LLMs through reading the first 2500 words of each paper. Lines smoothed using a 3-year moving average.

Fact 1: Theory meets data (not theory or data)

A plausible starting hypothesis regarding the impacts of the data revolution on the field of international trade is that the explosion of granular data, alongside the credibility revolution in applied econometrics, would have boosted the share of predominantly empirical papers within the field. Figure 1 presents the evolution of paper types across our three main categories over the sample period. We plot the share of papers across each category within each year, using a 3-year moving average to smooth out noise.

The striking patterns revealed in the figure do not support this hypothesis. Instead of growing, the share of primarily empirical work has remained relatively constant between 2000 and 2025 despite the revolutions in both data availability and empirical methodology. In contrast, the share of papers combining theoretical and empirical contributions has increased substantially, with this growth coming at the expense of primarily theoretical work. The share of primarily theoretical papers declined from roughly 40% during 2000-2004 to less than 10% during 2020-2025—a fourfold reduction. This is even true in absolute terms, with the average number of theoretical papers per year declining from about 40 to less than 10, despite the total number of trade papers published in these outlets over time growing substantially (see Appendix Figure 7 for plots of number of papers per year). Papers combining theory and empirics rose from roughly 25% to 60% of the trade literature over the same period. Meanwhile, primarily empirical papers have remained relatively stable, comprising around one third of the literature.

These patterns are consistent with the field increasingly demanding that both theoretical and empirical contributions be combined in a single paper, and from discussions with graduate students, these pressures are certainly present for those hoping to produce a strong job market paper. As we detail below, a theory paper that might have been published as a stand-alone contribution in the early 2000s now typically contains some form of empirical implementation, for example a quantitative model (the dominant subcategory, as we show below), tests of model predictions, or novel motivating evidence. A primarily empirical analysis that may have published well in the early 2000s is now more likely to contain a formal model, with a tight connection between that model and the empirics. This represents a fundamental shift in what constitutes a well-published contribution to the trade literature over time.

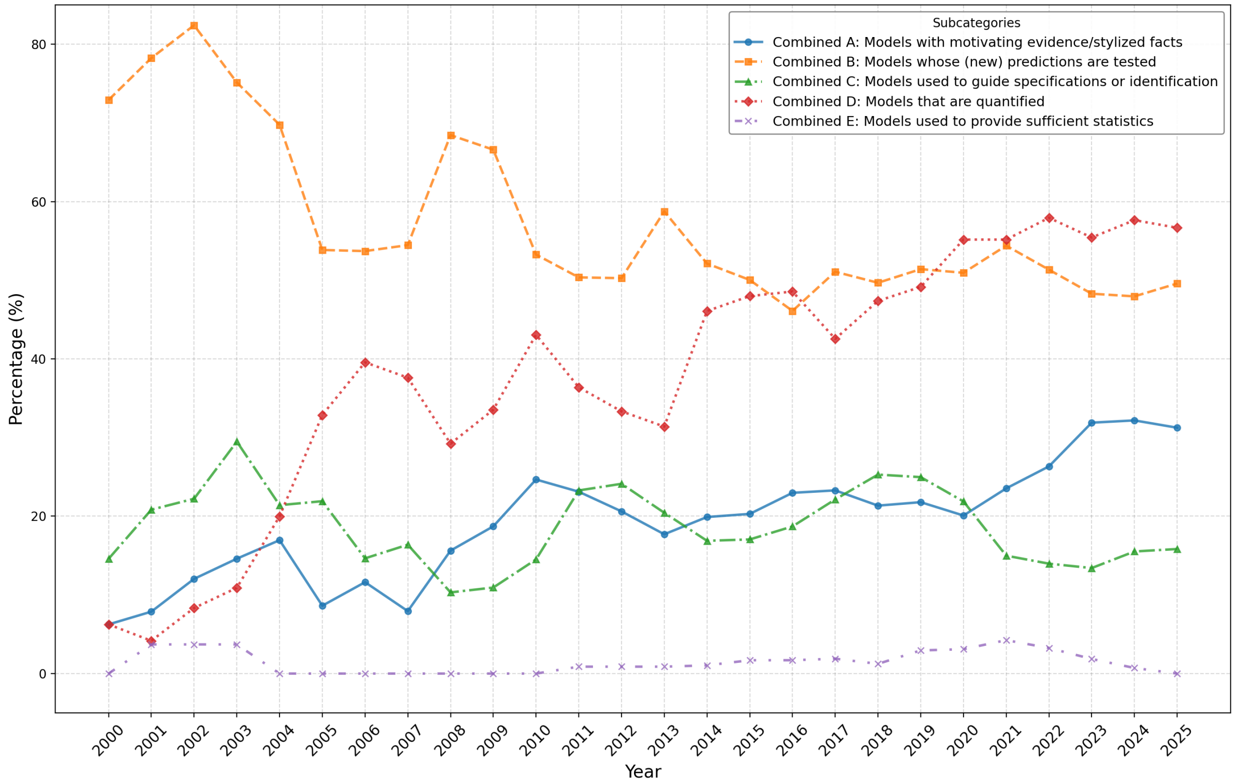
Interestingly, this pattern is even more pronounced in the top-five general-interest journals (AER, Econometrica, JPE, QJE, and REStud) compared to the leading field journal (JIE). Close to 60% of trade papers in top-five journals combine theory and empirics, versus roughly 40% in the JIE over this period. These generalist outlets also publish proportionately fewer primarily theoretical papers in international trade (15%) compared to the JIE (25%) (see Appendix Table 8). This differential across journal tiers is supportive of the hypothesis above that the field increasingly believes a theoretical contribution must be tightly linked to data to qualify as a major contribution, although we cannot fully dismiss the possibility that these patterns are supply driven.

Fact 2: The Turn to Quantification

Given the increase in papers that combine theory and empirics, an obvious next question is to explore the nature of these papers to discern which types of analysis have been behind this rise. Figure 2 plots the share of papers within this combined category that fall into the five subcategories delineated above: (Combined A) those where data are used as motivating evidence, (Combined B) those where data are used to test the model's predictions, (Combined C) those where the model is used to derive the empirical specifications and/or guide identification, (Combined D) those where data are used to quantify, and (Combined E) those where the model is used to derive a set of sufficient statistics that can be measured in the data. Each paper receives a weight of 1 for each subcategory it has been assigned to, with the y-axis showing the share of all combination papers that use that particular approach (with shares summing to more than 1 across subcategories as papers can appear in multiple subcategories). We find similar patterns when we assign an equal share of that paper to each subcategory that the LLM has assigned (so that the sum of weights across all subcategories equals the total number of papers in the combined theory and empirics category), as shown in Appendix Figure 10.

In the early years of our analysis, the rise in the share of papers that combine theory and empirics was mainly driven by papers that derived testable predictions from a new model and then turn to the data to test those predictions (Combined B). Approximately 70% of the combined papers published during the period 2000-2012 use this methodology. Only about 20% of papers used the data to quantify models for counterfactual analysis (Combined D). By the end of the second half of our sample, 2013-2025, these patterns had reversed. About 60% of papers used data to quantify models, while only around 50% of papers tested predictions.

Figure 2. : Evolution of Combined Theory-Empirics Subcategories 2000–2025



Notes: Figure 2 plots the share of papers in the field of international trade published in seven leading journals between 2000 and 2025 by subcategory within combined papers (that use both theory and empirics). Categorization done via LLMs through reading the first 2500 words of each paper. Lines are smoothed using a 3-year moving average. As papers can be assigned to multiple subcategories, percentages can sum to more than 100% in each year.

Thus, within this growing category of papers combining theory and empirics, the figure documents a shift from using data to *test* theories (Combined B) toward using data to *quantify* theories (Combined D). The rise of calibrated and/or estimated structural models—used to simulate counterfactual scenarios and decompose the contributions of different mechanisms—represents a methodological transformation in the interplay between theory and data.

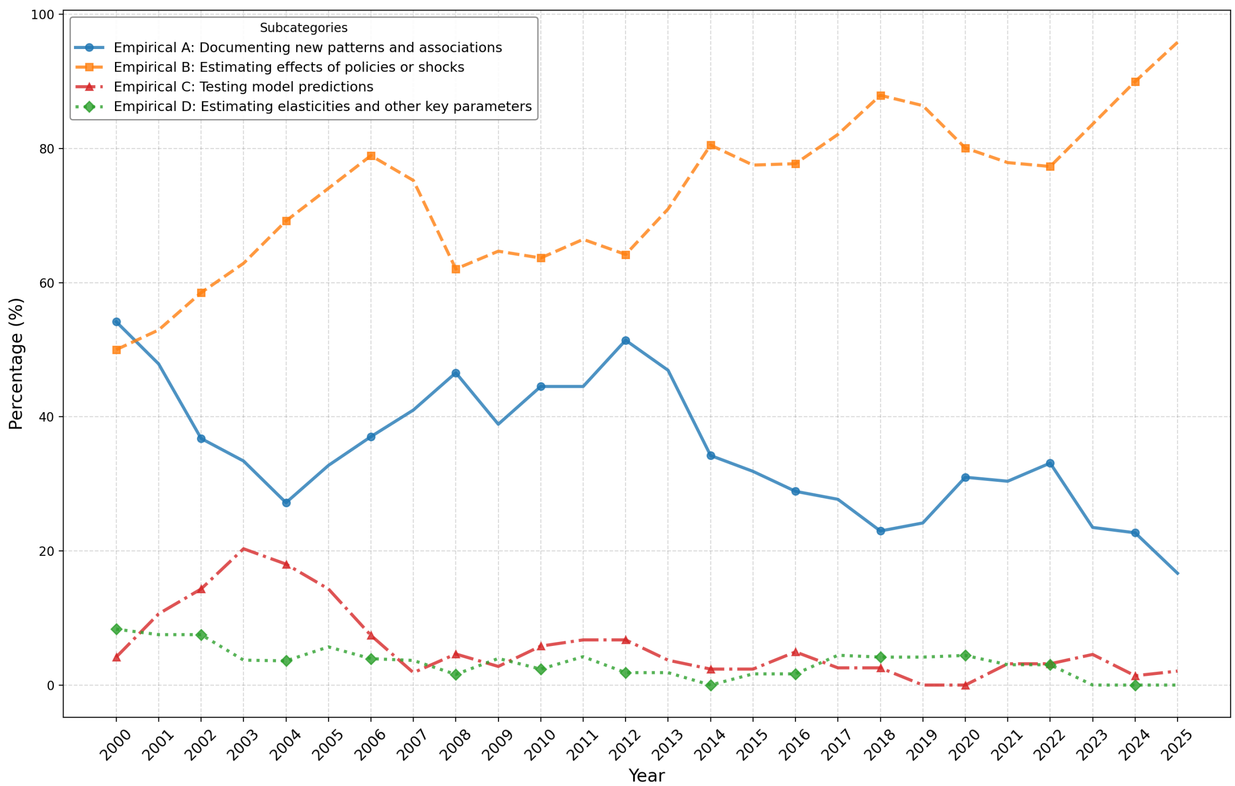
This shift has profound implications for the role of empirical evidence in trade research. In the testing paradigm, data serves primarily to validate or invalidate theoretical predictions; the theory is judged by whether its comparative statics match empirical patterns. In the quantification paradigm, data serves to discipline model parameters, enabling welfare calculations, counterfactual simulations, and decompositions that would not be possible from reduced-form analysis alone. Rather than searching for evidence that supports or invalidates a model, the question becomes to what extent the mechanisms captured by the model appear to be quantitatively important.

Multi-subcategory combinations are common and revealing. In the 2000s, the most frequent pairing is Combined A+B that present motivating evidence for a new theoretical mechanism before developing the model and testing its predictions (43% of the 28 papers that had multi-subcategory combinations). By the 2020s, each of the three most-frequently appearing combinations included quantitative work (A+B+D with 23%, B+D with 20% and A+D with 19% of the 84 papers).

Finally, the figure also shows the low occurrence of the two remaining subcategories: models primarily used to guide empirical specifications and identification, and models used to derive sufficient statistics from the data. While these approaches are powerful, and are perhaps the approaches that meld most closely with the primarily empirical literature, they have not been widely pursued.

B. The Credibility Revolution Within Empirical Trade

Figure 3. : Evolution of Empirical Paper Subcategories 2000–2025



Notes: Figure 3 plots the share of papers in the field of international trade published in seven leading journals between 2000 and 2025 by subcategory within primarily empirical papers. Categorization done via LLMs through reading the first 2500 words of each paper. Lines are smoothed using a 3-year moving average. As papers can be assigned to multiple subcategories, percentages can sum to more than 100% in each year.

Fact 3: The rise of causal inference

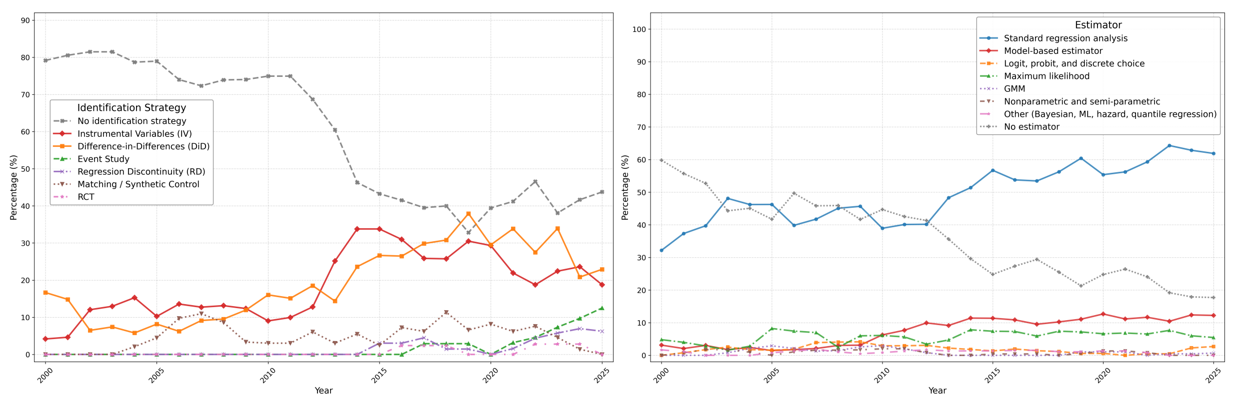
Over our period of study, the "credibility revolution" in applied economics unfolded. The adoption of the applied microeconomics toolkit—difference-in-differences, instrumental variables, regression discontinuity, randomized control trials, matching estimators etc.—and a deeper understanding of causal inference led to an explosion of reduced-form empirical work across economics.

In line with this context, we see that papers that estimate causal effects of policies, shocks, or interventions (Empirical B) have come to dominate the empirical landscape. Figure 3 shows the evolution of subcategories within primarily empirical contributions from 2000 to 2025. Again, each paper receives a weight of 1 for each subcategory it has been assigned to with qualitatively similar results assigning fractional weights in the appendix. In the early 2000s, more than half of empirical papers documented stylized facts (Empirical A) with a similar share for papers estimating impacts of polices or shocks where the techniques brought to the fore by the credibility revolution were most relevant (Empirical B). By the 2020s, these shares had shifted dramatically. More than 90% of empirical papers now estimate impacts of policies or shocks.

We conjecture that papers that once might have presented descriptive patterns or correlations now are expected to establish causality, naturally pushing towards estimating the impacts of policies or shocks (60 papers in our sample combine both approaches, documenting stylized facts before estimating impacts).

Fewer than 5% of primarily empirical papers test predictions of existing theoretical models (Empirical C) or estimate structural parameters of existing models (Empirical D). Although they never constituted a large share, if anything, these modes of inquiry have shrunk in relative importance over time (despite both being amenable to causal inference). This represents a notable departure from the 1980s and 1990s when, for example, gravity equation tests and factor content tests constituted major empirical literatures. Today, such work more commonly appears within combined theory-and-empirics papers rather than as standalone empirical contributions.

Figure 4. : Evolution of Identification Strategies and Estimators 2000–2025



(a) Evolution of Empirical Papers by Identification Strategy

(b) Evolution of All Papers by Estimator

Notes: Figure 4a plots the share of primarily empirical papers in the field of international trade published in seven leading journals between 2000 and 2025 by identification strategy. Figure 4b plots the share of all papers in the field of international trade published in the same journals by the type of estimator used in the paper. For both figures, categorization is done via LLMs through reading the first 4000 words of each paper. Lines are smoothed using a 3-year moving average. As papers can be assigned to multiple estimator and strategy types, percentages can sum to more than 100% in each year.

The rise of the credibility revolution is also apparent in Figure 4 that documents the use of different identification strategies among primarily empirical papers in the trade literature over time.[6] At the beginning of the sample in the early 2000s, over 80% of primarily emprical trade papers had no clear identification strategy. By 2025, that share had fallen to less than half. Among identification strategies, IV and difference in differences have been the dominant approaches, with roughly 30% of papers using each of these approaches.

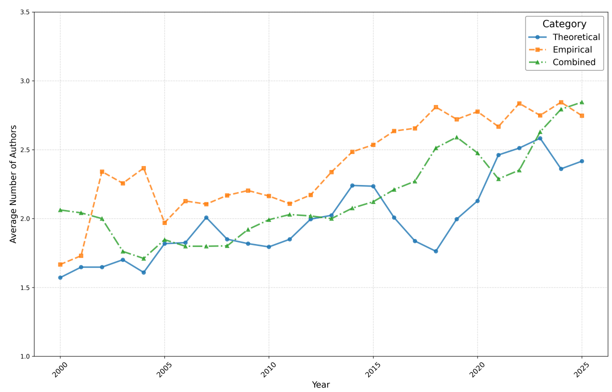
Finally, Figure 5b presents the evolution of the use of different types of estimators in the trade literature, grouping together closely-related types of estimators. To also capture the rise of model based estimators, we revert to our full corpus. The first point to note is the steady decline of papers that had no estimator at all, from around 60% to under 20% of papers. This reflects the decline of purely theoretical contributions as well as, to a lesser extent, the move away from

[6]In both cases we ask Claude Opus 4.5 to classify each paper's identification strategies and the primary estimators used. The full prompt is detailed in the Appendix.

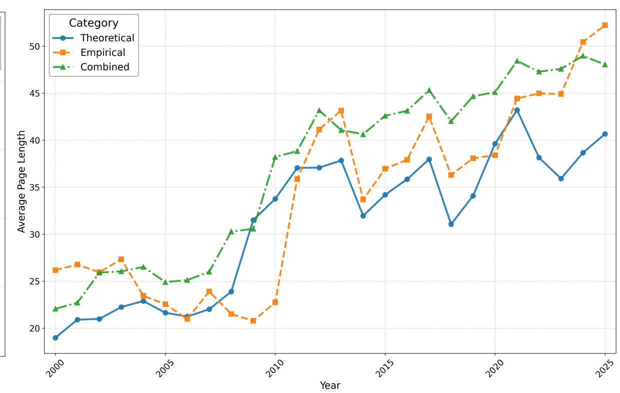
more descriptive papers among primarily empirical work, and the increasing use of parameter estimation (rather than traditional calibration) in quantitative work. In terms of which specific estimators researchers turned to, the share of papers using standard regression-based estimators such as OLS, two stage least squares, or linear fixed effects estimators rose from around 30% to 60% of all papers. Consistent with this rise, the credibility revolution elevated these standard estimators which became the workhorses of causal estimation, in part because of their transparency. Reflecting the quantitative turn in international trade, the share of papers employing model-based estimators (fitting data moments to a model to estimate parameters through minimum-distance estimators, simulated method of moments or indirect inference) tripled from 4% to 12% during the sample period.

C. The Changing Paper Production Process

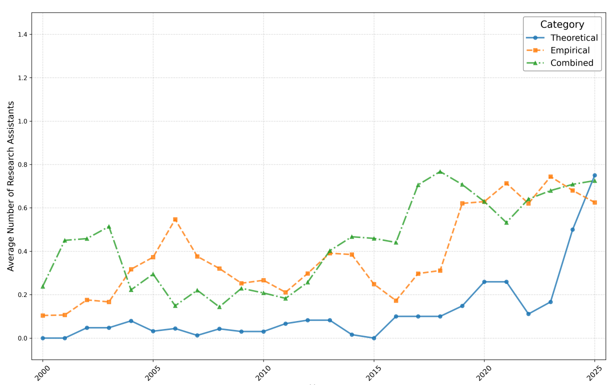
Figure 5. : Evolution of Counts of Authors, Pages, and Research Assistants 2000-2025



(a) Average Number of Coauthors 2000-2025



(b) Average Page Length 2000-2025



(c) Average Number of Research Assistants

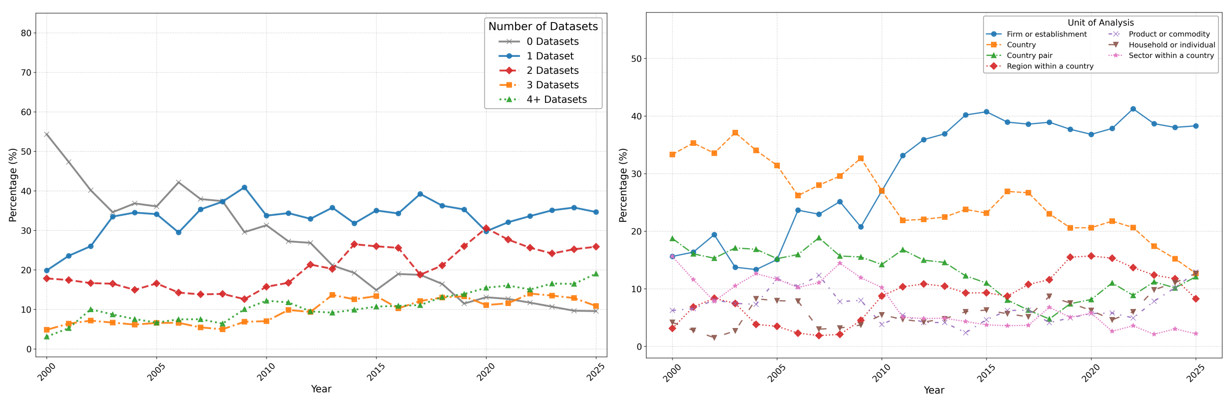
Notes: Figure 5a plots the average number of authors per paper among papers published in seven leading journals within international trade by paper category. Figure 5b reproduces the plot but with the average page length of papers published while Figure 5c plots the number of research assistants acknowledged. Lines are smoothed using a 3-year moving average.

Fact 4: More coauthors, more research assistants, more pages, more datasets

In line with these changes in how empirical and theoretical contributions are published in international trade, we also document changes in the paper production process over this period. Figure 6a, shows the average number of authors per paper and Figure 6b the average page count. The latter more than doubled for all three broad categories from 20-25 pages on average to more than 50 over the past 25 years. The number of authors increased substantially, from about 1.7 to 2.8 authors per paper. In Figure 6c, we also find a steady increase in the fraction of papers in international trade that mention the support of research assistants—as captured by mentions in the disclaimer notes on the title page.

While these metrics almost certainly provide an incomplete account how the research process has changed over time, the secular rise in all these objects supports the notion that papers in international trade are increasingly combining a broader set of methods—most notable in the increase in papers combining theory with empirics—which requires more inputs and more diverse skillsets. Having said this, we also see a steady increase in page length, number of authors, and RA usage among primarily theoretical and primarily empirical papers, suggesting that other forces beyond the need for more diverse skillsets are operating over this period.

Figure 6. : Evolution of Data Usage 2000--2025



(a) Evolution of Number of Datasets Per Paper 2000-2025

(b) Evolution of the Unit of Analysis

Notes: Figure 6a plots the share of papers by the number of datasets used in each paper (0, 1, 2, 3, or 4+ datasets) for all papers. Lines are smoothed using a 3-year moving average. Percentages sum to 100% in each year. Figure 6b plots the share of papers by the unit of analysis the authors focus on (for the subsample of all papers with at least one dataset). In Figure 6b, percentages can exceed 100% as papers can use multiple units of analysis.

The number of datasets per paper and the type of data used in the literature have also changed over our sample period. Figure 6a plots the share of papers using 0, 1, 2, 3, or 4 or more datasets over time for all papers in our corpus.[7] In line with the decline of primarily theoretical work in international trade, the fraction of papers using no datasets declined from more than 50% in the early 2000s to about 10% by the end of the sample. In contrast, the steepest increase was in papers using 4 or more datasets (rising from close to 0% to about 20%). Figure 6b documents the type of data used in papers with at least one dataset, distinguishing between the unit of analysis: countries, country pairs, sectors, regions within a country, firms or establishments, households or individuals, etc. In the early 2000s almost 40% of papers used country-level data. This fell to under 15% by 2025, with steep increases in the share of papers using firm- or establishment-level microdata. The rise of spatial economics is also visible in the doubling of the share of papers using data across regions within a country (for example, states, counties, or municipalities).

[7]We used Claude Opus 4.5 to determine the number of datasets used in the paper from the first 4000 words of text in the paper, rechecking with the same LLM when a paper had more than five datasets.

Taken together, these findings paint a picture of a field that has fundamentally transformed how it uses data. The data revolution in international trade has not meant more empirical work at the expense of theory. Rather, it has been accompanied by a new relationship between theory and empirics—one that was not obviously the result of greater data availability alone.

A naive observer in 2000, told that, in the broader context of the credibility revolution in economics, trade economists would soon also have widespread access to transaction-level trade data, matched employer-employee records, economic census microdata, rich administrative records including from local governments within countries and satellite imagery, might reasonably have predicted an explosion in well-published papers that were primarily empirical—more careful documentation of new phenomena, more rigorous testing of existing theories, and more precise estimation of key parameters.

Such an observer would have been only partially correct. What actually occurred was a pronounced shift toward model calibration and counterfactual analysis, at the expense of pure theory and likely at the expense of pure empirics whose share of the literature remained flat despite these increasing opportunities. Even among papers combining both theory and empirics, more granular data and more sophisticated econometric tools could have led the field toward more rigorous testing of theoretical predictions—using quasi-experimental variation to adjudicate between competing models, for example. Instead, the field moved toward a paradigm where models are calibrated to match moments and then used for counterfactual simulations. Testing, when it occurs, often involves showing that a model matches untargeted moments rather than that it survives attempts at falsification.

An obvious question is whether or not these developments are unique to the field of international trade. While a full analysis of papers published across all fields is beyond the scope of this review, recent work by Goldsmith-Pinkham (2024) suggests these developments were not universal. Reviewing methods mentioned in NBER working papers (i.e. papers by leading US-based scholars in applied fields and macroeconomics), he finds that papers released by members of the International Trade and Investment research group are less likely than most to mention identification (11th out of 19 groups) or experimental or quasi-experimental methods (15th of 19), and more likely to mention structural estimation (3rd of 19). One feature that may explain these patterns is that many questions in the trade field are inherently general equilibrium, which makes them less amenable to experimental and quasi-experimental methods than questions in some other fields. This focus may also help to explain why research in international trade has shifted towards combining theory and data—for example a model may be necessary to measure changes in the intercept that cannot be identified by variation across firms, industries, or locations.

III. Using data and theory more effectively

The analysis above reveals a dramatic change in the types of papers trade economists write. Historically, trade economists used to specialize more. Some would write pure theory, some would do pure empirical work, while others would combine less path-breaking analytical modeling with sign tests of key comparative statics that those models produced. Now researchers are often expected to do all of these in the same quantitative paper, particularly students on the job market. Given this dramatic change of direction for the field that has accompanied the data revolution, in this section we assess its implications and consider two dimensions along which we believe theory and data can be combined more effectively.

In the hands of the most skilled practitioners, a tight interplay between theory and data can be hugely insightful. This interplay may take several forms. The model can clarify a puzzling or difficult-to-interpret empirical result, or provide a specification that uncovers a structural parameter which is more externally valid (for example, in the presence of SUTVA violations due to spillovers or when measured variables need to be appropriately adjusted to reveal the causal relationship of interest). Alternatively, a model may allow the authors to answer indirect causal questions that cannot be revealed in a reduced form manner. The data may also reject features of the model over which the author had strong priors and lead them in a more fruitful direction, or show that a compelling mechanism was not actually the primary driver of an outcome that had been attributed to it. But there is also the possibility that we may limit our collective understanding by establishing the quantitative trade paradigm as the default mode of inquiry in the field, a default the LLM analysis above suggests may be happening. It is perhaps revealing that the clearest examples of major new insights regarding how trade impacts the world that have been uncovered over the last 25 years have come either from work that is mostly theoretical (e.g. the Melitz model of firm heterogeneity and its extensions to studying the distributional implications of trade), work that is mostly empirical (e.g. the local labor market impacts of trade or the pass-through of tariffs to border prices), or events in the world that the academic trade community was arguably as surprised by as policymakers were (e.g. supply chain resilience or geoeconomics).

There are several possibilities here. One is simply that many foundational questions were already answered in the previous decades that brought forth new trade theory, strategic trade policy and protection for sale, and economic geography. Another possibility is that the abundance of granular data encourages incremental research at the expense of more fundamental exploration. For example, to make the most of some exciting and highly granular new data set, the author now also requires a quantitative model that features heterogeneity on new dimensions but those additions may make the model less transparent and therefore harder to extract the core economics of the question.[8] It is not uncommon for the reader to have little idea what fundamental economic forces, implicit in the model, drive the key quantification result (e.g., what are the key inefficiencies being ameliorated such that gains from trade are amplified?).

A third possibility is that by trying to write a paper that both explores a novel dataset and goes on to combine that data with a frontier quantitative model, less space is allocated to either the empirical data exploration or to the interplay of economic mechanisms than were the empirical and theoretical elements split across two papers. This would be problematic if paradigm-shifting insights come from exploration at the margins or from research skills acquired through specialization.

Within categories, the LLM analysis also uncovered substantial changes in the types of papers that trade economists publish. Within empirical work, there has been a pronounced shift to estimating the impacts of policies and shocks. This is perhaps not surprising given that the credibility revolution in applied economics emphasizes the use of experiments and quasi-experiments. The revolution's origins in the 1980s and 1990s were closely tied to dissatisfaction with existing methods for impact evaluation (see, for example, the work of LaLonde (1986)). While we believe this is a favorable development, the push toward estimating causal impacts in primarily empirical work may have come at the expense of more exploratory empirical work, for example, survey evidence on reactions to trade policies, or studying important economic relationships that are less amenable to causal identification, such as globalization's gradual effects on national institutions and property-rights enforcement. We also emphasize above that, despite the explosion of work evaluating policy shocks, there still remains a shortage of credible estimates of some of the key elasticities and parameters used in the field.

Finally, within work that combines theory and empirics, quantitative approaches have caught up with the once dominant category of testing a model's predictions using observable comparative statics. We now turn to discussing several aspects of these developments, as well as ways we believe these tools can be used more effectively within the trade field.

[8] Sometimes granularity may even limit the ability to answer a question, as pointed out in Dingel and Tintelnot (2023) where aggregated data can provide more reliable answers than taking the granular data at face value.

Different goals, different approaches

There are many fruitful approaches to research in the field of international trade. These include: exploratory data work that uses the growing wealth of granular information to document new, and maybe ex ante puzzling, patterns that are clearly important but the author may not yet have a tight explanation for; reduced-form empirical work with a focus on identification that documents the effects of policy shocks along a comprehensive vector of different outcomes of interest that may be too numerous to all be rationalized by a single quantitative model; new ideas most clearly laid out through elegant applied theory that is far too simplified to merit quantification; or important ideas for which the data required to understand their quantitative importance is not available. One may be concerned, as we are, that the extent to which the field has shifted away from more focused and diverse types of analysis such as these and toward the combination of theoretical and empirical contributions can both limit what we can learn and reduce the gains from specialization. Of course, the growing number of coauthors as well as the increase in pages per paper suggest that there may be greater specialization within papers, but the norms (and journal limits) regarding paper length are sufficiently strong to generate meaningful tradeoffs.

That said, the combination of theory and empirics offers the promise of improving on what either can offer alone. As with work across broad categories, we also believe that there are gains to specialization within quantitative work, albeit of a different form. As the analysis above highlights, researchers are increasingly turning to quantitative estimation. This development has given rise to what may be described as a cookie-cutter framework for writing papers in the trade field. First, one motivates the paper with a policy question of interest, then writes down a model with relevant features of the economy, estimates or calibrates that model with data, and finally performs counterfactuals to answer the policy question. Within this framework, there are (at least) three very different goals that are often conflated. We hypothesize that our understanding of the world would grow faster if there was more freedom to deviate from this framework where it holds back our ability to communicate the knowledge generated by the research process.

The first goal is the one that the paper's introduction typically opens with—credibly answering the policy question of interest. If providing accurate numbers to policy-makers was the primary goal, one would want the identifying variation to be as close as possible to the policy question, care a lot about bias and standard errors, and try to include all consequential model features, however inelegant they may be (and potentially strip out less consequential but standard ones). The work done by central banks predicting the impacts of different monetary policy choices as inputs into monetary policy decision making would be an example of how this type of work would be done if providing credible numbers was paramount. These models are good for prediction but less good for understanding the economic mechanisms in the world, which is why trade economists have tended to shy away from them just as they shied away from the Computable General Equilibrium models that are still used by policymakers today to make predictions about the impacts of trade reforms and such like (see, for example, the book by Hertel (1997)). While many trade economists see these models as inscrutable black boxes, that view is likely shared by academic macroeconomists viewing the central bank models that provide policymakers with likely paths of inflation caused by monetary or fiscal policy choices. We suspect few academic macroeconomists would claim their more analytically tractable models make better predictions of the impacts of policy choices. Instead, these stripped down academic models provide specific insights that filter through to the models used by central banks.

It is difficult to provide compelling examples in trade of papers that are very concerned with obtaining accurate and believable counterfactuals that directly inform policy choices. One of the best recent examples of such a paper comes from Fajgelbaum et al. (2020) who analyze the impacts of Trump's first-term tariffs by combining rich trade data, convincing event study designs, and a frontier trade model to calculate welfare impacts. In this case, the policy interest is immediate, and we use this paper as an example because it is an excellent and highly-informative paper. However even here, much of the effort in the analysis does not go to generating plausible estimates of the welfare impacts of the tariffs. As is well known, the aggregate welfare impacts of the tariffs hinge critically on the terms of trade impacts which, in the paper's model, are not identified by price changes across foreign varieties that is the centerpiece of the empirical analysis but instead the sector-level elasticities of US supply that only appear in the appendix and are less cleanly identified. Thus, the paper finds meaningful terms of trade improvements due to the tariff war despite the headline empirical finding that Chinese firms did not lower their export prices. If we wanted to put accurate numbers on the welfare effects of the Trump tariffs, we should focus our effort on identifying the key elasticity that pins down those welfare impacts or on rigorous model testing—two points we return to below (as we note there, Adão, Costinot and Donaldson (2025) test this specific model and find it performs well). Alternatively, we may want a paper centered on well-identified causal effects and exploring how they relate to important policy questions beyond the welfare impacts.

A second goal of quantitative analysis is to use the back and forth with the data to discipline the model structure, learn something about how forces interact with each other, and arrive at a more sensible model of the world. The authors use their priors on the direction and size of various impacts to reject certain model assumptions or approaches. For example, if modeling phenomenon Y by assuming X rather than Z leads to price changes that are an order of magnitude larger than the authors' priors (or estimates from other work), the model should assume Z. This type of analysis often contains explicit propositions and has close connections to an older generation of papers by applied theorists that used relationships in the data in a similar manner—to motivate a theory or guide assumptions. Such an approach is hugely useful when it allows us to rule out certain model structures, mechanisms, or functional forms that are at odds with the data. These may be the most valuable insights from the paper. However, these insights are directly at odds with being able to credibly answer a policy question. This back and forth with data means answers will be biased towards the authors' priors. Additionally, there is little benefit to sharing what has been learned about how certain assumptions generate implausible answers—such an admission would suggest the quantitative answers are not robust to alternative assumptions and cannot be trusted for answering a policy question. Of course, that is not to say that this type of approach is not policy relevant. However, unlike papers that aim to provide credible counterfactuals for actual or potential policies, here the path is more indirect, with the hope being that the paper convinces policymakers and other economists that the mechanisms highlighted are important and should be included in a more comprehensive assessment of particular policy choices.

By way of illustration, it is informative to provide an example from our own work of the value quantitative modeling can bring. In a recent paper by Atkin, Costinot and Fukui (2021), we wrote down an analytically tractable model that shows that trade helps all countries climb the ladder of development by shifting them into their most complex sectors where spillovers are highest. Quantifying a more general model that had more flexibility to match the data revealed the opposite result. After confirming this was not a coding error, the quantification led us to discover that more complex goods are produced by many more countries than simpler goods and so trade pushes you away from high spillover sectors. This was ruled out by our technological assumptions as it hadn't occurred to us as a possibility given the literature on country capability popularized by Ricardo Hausman and coauthors. We make no pretense that our quantification provides reasonable estimates of the dynamic gains from trade—instead it revealed to us a feature of the data that is at odds with the ladder of development metaphor commonly discussed in policy circles. Consistent with the cookie cutter approach to quantitative work we highlight above, the publication process required us to pretend we knew this all along and start the paper with the more general model where anything can happen and then proceed to quantify it.

A third goal of quantitative analysis is to facilitate the exploration of models not amenable to analytical solutions (and do so in relevant parameter space). Here again the straitjacket of organizing the paper around answering a policy question can limit what we learn. If the explicit goal was to understand a complicated model in a reasonable parameter space, the analysis would be more exhaustive and convincing as it would not need to focus on one single set of parameters that are used for the policy counterfactual. The hypothesis being examined computationally could be clearly stated, and how the answers depend on a full range of sensible parameters and secondary model features explored. Of course, many papers do this to some degree but the fact that it often appears as a robustness exercise to the main numbers constrains how this information is discussed and presented and what the reader learns from it. There is a long history of examining models through simulations and we believe that more is learned when practitioners are up front about this goal.

To summarize, in terms of broad categories of research, moving away from a diverse and more specialized set of approaches and methods gives rise to concerns about limiting discovery. Furthermore, within quantitative work that combines theory and data, more could be learned and at a faster rate if we moved away from the cookie-cutter framework of motivating the paper with some policy counterfactual of interest which needs to be quantified and instead embraced more fully the quantitative theorizing that is often the central contribution rather than credible counterfactual numbers—allowing the authors to better communicate the knowledge actually created through the paper production process.

Improving Policy Counterfactuals

We have just advocated for moving away from a cookie-cutter approach to quantitative research where almost all papers are set up in pursuit of quantifying a counterfactual of interest—whether or not the efforts of the authors are primarily devoted to producing credible answers to that counterfactual. However, at the same time policy counterfactuals that inform real-world questions (the first goal in quantitative papers discussed above) remain incredibly important. For papers where this is the primary goal, we believe that the credibility of these counterfactuals can be substantially improved through more model testing, more credible and relevant moments, and more structured sensitivity analysis.

An older literature pioneered by Kehoe, Polo and Sancho (1995), Kehoe (2005) and extended in Kehoe, Pujolas and Rossbach (2017) assessed the predictive performance of calibrated trade models by comparing their counterfactual predictions to observed outcomes following actual policy changes. The sobering finding from this work is that these models often failed to predict which sectors would experience the largest changes in trade flows, even when the direction of aggregate changes was correct. A newer generation of more rigorous tests has emerged. Adão, Costinot and Donaldson (2025) develop a framework for testing whether quantitative trade models can correctly predict the effects of observed policy changes—in their application, the Trump tariffs—by comparing model-implied cross-sectional predictions to those estimated using causal identification strategies. This approach is closely related to work where quantitative models are assessed against quasi-experimental variation, such as Ahlfeldt et al. (2015) who use the impact of the fall of the Berlin Wall on different parts of Berlin, or Donaldson (2018) who shows how much of the reduced-form effect of new railroad lines in India is captured by his structural trade model. Less formally, many papers now assess model fit using "untargeted moments"—data patterns the model was not calibrated to match—though the choice of which moments to target versus test against is often left unclear.

We believe more testing would strengthen the literature. There should be greater use of approaches such as comparing model predictions to causally-identified impacts of shocks, rather than simply assuming that because the model matches targeted moments it will correctly predict counterfactuals. Pre-specifying which moments will be used for estimation versus assessment would help address concerns about specification search. Some researchers have also begun to preregister structural analyses, as in the recent paper by Walker et al. (2024), though enforcement is more difficult than with randomized trials where data is generated after registration. Allen (2014) provides an interesting case where new data became available after the paper was written, allowing the model's predictions to be assessed against information that could not have influenced model specification—though this approach remains subject to the concern addressed by Adão, Costinot and Donaldson (2025) that unobserved shocks may vary over time.

A related concern is whether the moments used to estimate key model parameters come from variation that is relevant to the policy question at hand. Often the welfare impacts within complicated models boil down to one key elasticity and observable levels of exposure (as will necessarily be the case in efficient models where to assess the impacts of large shock one can apply Hulten's theorem repeatedly alongside changes in exposure that are pinned down by this key elasticity). Thus, a model's counterfactual predictions will be most credible when the identifying variation used to estimate these elasticities or parameters closely resembles the policy change being analyzed. For example, if the question is how changes in trade affect economic growth, estimating the key elasticity from cross-sectional variation in firm size distributions is likely to yield predictions that are less robust to alternative specifications than if they came from variation in trade exposure itself.

Bergquist, McIntosh and Startz (2024) provide a valuable example. Their paper develops a structural model of agricultural trade in which key parameters could, in principle, be estimated from observational data. Instead, they use a large-scale randomized controlled trial to estimate search frictions. This approach provides confidence that the model's parameterization will perform well at scale and reveals surprises—the intervention helped large farmers more than small ones, contrary to their priors.

While using true experimental variation is challenging if not impossible in many contexts, we also advocate for the greater use of natural experiments in providing credibly-identified moments for the estimation of quantitative models, as for example in the work of Ahlfeldt et al. (2015) or Faber and Gaubert (2019) discussed above. Goldsmith-Pinkham (2024) provides suggestive evidence that the trade field lags other fields along this dimension, documenting that out of NBER working papers that mention structural estimation, the share mentioning experimental or quasi-experimental methods is unusually low for the International Trade and Investment group (only 20%, the 5th lowest out of 19 NBER working groups).

Finally, the field would benefit from adopting tools for assessing sensitivity, including sensitivity to alternative functional forms and moment selections. Most directly, it is still not the norm to use estimation approaches within quantitative models that allow for standard errors to be placed on model counterfactuals. As documented in Figure 5b, there has been progress here with a marked reduction over time in papers with some empirical component that do not mention any kind of estimator. However, few papers attempt to quantify the uncertainty over the model outputs coming from the fact that parameters are estimated rather than known. Recent work by Andrews, Gentzkow and Shapiro (2017) on the sensitivity of model-based counterfactuals goes even further and acknowledges that forces outside the model may be driving data moments targeted inside the model, and develops an approach using different priors on the change in the moments used in the quantification (due to confounding forces) to assess the sensitivity of the counterfactual results to such alternative assumptions. Related work by Andrews, Gentzkow and Shapiro (2020) develops a method to quantify the 'informativeness' of the reported motivating evidence and descriptive statistics in driving model-based counterfactual results—to what extent can the presented evidence reduce the 'worst-case bias' from violating the model's assumptions.[9]

[9] More recent contributions in this space include those by Andrews et al. (2023) and Andrews and Shapiro (2024).

IV. Future directions

The data revolution has not run its course and we expect it to continue to influence the direction of the trade field. New data sources will continue to be unlocked and technological innovations will mean measurement becomes ever more granular and timely. We discuss several emerging data sources with the potential to reshape trade research in the coming years. First, global value chain data from within large firms is becoming more available as companies adopt software to track internal supply chains and implement traceability systems. Second, smart contracts and blockchain technologies are improving the documentation of commercial terms and enabling researchers to study contractual arrangements at scale. Third, text and audio data—from firms' quarterly reports and earnings calls to WTO negotiation records and tariff code documentation—can now be processed using natural language methods to extract quantitative measures of policy uncertainty, trade barriers, and firm expectations. Fourth, tracking data from ships, trucks, aircraft, and even individual containers provides real-time information on trade logistics. Fifth, satellite and remote sensing data continues to expand, enabling measurement of economic activity at remarkable granularity—from monitoring oil tanker movements (as in the work of Fernández-Villaverde et al. (2025)) to detecting inventory buildups outside factories (as studied by Dai, Chen and Zuo (2023)). Sixth, the digitization of both firm and individual level transactions is generating new administrative data on flows of payments and purchases at an unprecedented scale and granularity (see, for example, Atkin et al. (2025)'s work in Chile). Finally, barcode-level data with country-of-origin information, the spread of mobile payment systems around the world, and online price scrapers are improving measurement of prices and product characteristics (as in the work of Cavallo et al. (2021), Bai, Jaccard and Stumpner (2025) and Argente, Méndez and Van Patten (2025)).

These and other new data sources will create opportunities but also challenges. Some may help address the issues raised in Section III—for example, real-time tracking data could improve our ability to test model predictions against observed responses to shocks using event-studies and other empirical designs. However, the abundance of new data at ever more granular levels may also tempt researchers toward increasingly complex models whose mechanisms become difficult to interpret, reinforcing some of the concerns discussed above.

How will artificial intelligence (AI) affect how data and theory are used in international trade? Several of the concerns raised above—e.g., that rich data can motivate complicated models whose core mechanisms are hard to uncover or test—are likely to grow more relevant as AI transforms research practices. We anticipate that AI capabilities will substantially reduce the costs of quantitative counterfactual analysis. Current AI systems can already assist with writing code, making incremental extensions to frontier models, pulling together datasets, and even estimating models or simulating counterfactuals. The intuition is that it is easier to automate the coding and calibration of a structural model than to assess that model's validity on the dimensions most relevant for policy-making. Advances in AI and machine learning may also improve structural model methodologies. For example, Chen et al. (2026) show that initializing a neural network on data generated by a structural model and then retraining it on empirical data can help the network adjust for misspecification in the structural model, improving both fit vis-a-vis the structural model and out-of-sample performance vis-a-vis a pure machine-learning model.

AI may also transform reduced-form empirical research in powerful ways. For example, it may soon be straightforward for an LLM to mine ninety years of the Federal Register to identify natural experiments related to industry-specific trade policy (validating parallel trends and other assumptions as part of the process), then combine this corpus with detailed trade datasets to explore which of these experiments yield clean event-study figures. While this approach poses challenges regarding false positives, these are not fundamentally different from the generic overfitting concerns that accompany machine learning methods more broadly—challenges on which substantial methodological progress has been made (e.g., see the work of Mullainathan and Spiess (2017)).

The bottom line is that we expect substantial increases in the quantity of research produced, but at the same time many of the issues highlighted in this review will remain pertinent—and may become even more pronounced as the costs of producing both reduced form and quantitative work decline substantially.

REFERENCES

Adao, Rodrigo, Costas Arkolakis, and Federico Esposito. 2024. "General Equilibrium Effects in Space: Theory and Measurement." *American Economic Journal: Macroeconomics.* Conditionally accepted.

Adao, Rodrigo, Paul Carrillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz. 2020. "Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence."

Adão, Rodrigo, Arnaud Costinot, and Dave Donaldson. 2025. "Putting Quantitative Models to the Test: An Application to the U.S.-China Trade War*." *The Quarterly Journal of Economics*, 140(2): 1471–1524.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence From the Berlin Wall." *Econometrica*, 83(6): 2127–2189.

Alfaro, Laura, and Andrew Charlton. 2009. "Intra-industry Foreign Direct Investment." *American Economic Review*, 99(5): 2096–2119.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P Vasquez. 2022. "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages*." *The Quarterly Journal of Economics*, 137(3): 1495–1552.

Allen, Treb. 2014. "Information Frictions in Trade." *Econometrica*, 82(6): 2041–2083.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy *." *The Quarterly Journal of Economics*, 129(3): 1085–1140.

Amiti, Mary, Stephen J. Redding, and David E. Weinstein. 2019. "The Impact of the 2018 Tariffs on Prices and Welfare." *Journal of Economic Perspectives*, 33(4): 187–210.

Andrews, Isaiah, and Jesse M Shapiro. 2024. "Communicating scientific uncertainty via approximate posteriors." National Bureau of Economic Research.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments*." *The Quarterly Journal of Economics*, 132(4): 1553–1592.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M Shapiro. 2020. "On the informativeness of descriptive statistics for structural estimates." *Econometrica*, 88(6): 2231–2258.

Andrews, Isaiah, Nano Barahona, Matthew Gentzkow, Ashesh Rambachan, and Jesse M. Shapiro. 2023. "Structural Estimation Under Misspecification: Theory and Implications for Practice."

Antras, Pol. 2003. "Firms, Contract, and Trade Structure." *The Quarterly Journal of Economics*, 118(4): 1375–1418.

Antràs, Pol, Teresa C. Fort, and Felix Tintelnot. 2017. "The Margins of Global Sourcing: Theory and Evidence from US Firms." *American Economic Review*, 107(9): 2514–2564.

Argente, David, Esteban Méndez, and Diana Van Patten. 2025. "Cross-border product adoption: Individual imports, migrant networks, and domestic retailers." *Yale University Working Paper.*

Arkolakis, Costas, Arnaud Costinot, and Andres Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94–130.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andres Rodríguez-Clare. 2019. "The Elusive Pro-Competitive Effects of Trade." *Review of Economic Studies.*

Arkolakis, Costas, Fabian Eckert, and Rowan Shi. 2023. "Combinatorial Discrete Choice: Theory and Application to Multinational Production." *Review of Economic Studies.* Forthcoming.

Arkolakis, Costas, Federico Huneeus, and Yuhei Miyauchi. 2025. "Production Network Formation, Trade, and Welfare." Working Paper.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2014. "Vertical Integration and Input Flows." *American Economic Review*, 104(4): 1120–1148.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-market, trade costs, and international relative prices." *American Economic Review*, 98(5): 1998–2031.

Atkin, David, Amit Khandelwal, and Adam Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment." *Quarterly Journal of Economics*, 132(2): 1101–1164.

Atkin, David, Arnaud Costinot, and Masao Fukui. 2021. "Globalization and the Ladder of Development: Pushed to the Top or Held at the Bottom?"

Atkin, David, Baptiste Bernadac, Dave Donaldson, Tishara Garg, and Federico Huneeus. 2025. "The Incidence of Distortions." Banco Central de Chile.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2018. "Retail Globalization and Household Welfare: Evidence from Mexico." *Journal of Political Economy*, 126(1): 1–73.

Atkin, David, Joaquin Blaum, Pablo Fajgelbaum, and Augusto Ospital. 2024. "Trade Barriers and Market Power: Evidence from Argentina's Discretionary Import Restrictions."

Autor, David, David Dorn, and Gordon H Hanson. 2021. "On the Persistence of the China Shock." National Bureau of Economic Research Working Paper 29401.

Autor, David, David Dorn, Gordon H. Hanson, Maggie R. Jones, and Bradley Setzler. 2025. "Places versus People: The Ins and Outs of Labor Market Adjustment to Globalization."

Autor, D., D. Dorn, and G. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103: 2121–2168.

Bai, Liang, Torsten Jaccard, and Sebastian Stumpner. 2025. "Tariff Pass-through into Retail Prices: Measurement, Replacement, and Shrink-flation."

Bartelme, Dominick, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2025. "The Textbook Case for Industrial Policy: Theory Meets Data." *Journal of Political Economy*, 133(5): 1527–1573.

Bergquist, Lauren, Craig McIntosh, and Meredith Startz. 2024. "Search Costs, Intermediation, and Trade: Experimental Evidence from Ugandan Agricultural Markets."

Bernard, Andew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *American Economic Review*, 93(4): 1268–1290.

Bernard, Andrew B., and Andreas Moxnes. 2018. "Networks and Trade." In *Annual Review of Economics.* Vol. 10, 65–85.

Bernard, Andrew B., and J. Bradford Jensen. 1995. "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987." *Brookings Papers: Microeconomics*, 67–119.

Bernard, Andrew B., and J. Bradford Jensen. 1999. "Exceptional Exporter Performance: Cause, Effect, or Both." *Journal of International Economics*, 47: 1–25.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2012. "The Empirics of Firm Heterogeneity and International Trade." *Annual Review of Economics*, 4: 283–313.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2011. "Multi-Product Firms and Trade Liberalization." *Quarterly Journal of Economics*, 126(3): 1271–1318.

Bernhofen, Daniel, and John Brown. 2004. "A Direct Test of the Theory of Comparative Advantage: The Case of Japan." *The Journal of Political Economy*, 112(1): pp. 48–67.

Blaum, Joaquin, Claire Lelarge, and Michael Peters. 2018. "The Gains from Input Trade with Heterogeneous Importers." *American Economic Journal: Macroeconomics*, 10(4): 77–127.

Boehm, Christoph, Andrei A. Levchenko, Nitya Pandalai-Nayar, and Hiroshi Toma. 2024. "Dynamic Models, New Gains from Trade?" Issue: 32565 Series: Working Paper Series Type: Working Paper _eprint: 32565.

Boehm, Christoph E., Andrei A. Levchenko, and Nitya Pandalai-Nayar. 2023. "The Long and Short (Run) of Trade Elasticities." *American Economic Review*, 113(4): 861–905.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies*, 89(1): 181–213.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2025. "A Practical Guide to Shift-Share Instruments." *Journal of Economic Perspectives*, 39(1): 181–204.

Borusyak, Kirill, Rafael Dix-Carneiro, and Brian Kovak. 2023. "Understanding Migration Responses to Local Shocks." Duke University.

Bowen, Harry P., Edward E. Leamer, and Leo Sveikauskas. 1987. "Multicountry, Multifactor Tests of the Factor Abundance Theory." *American Economic Review*, 77(5): 791–809.

Bryan, Gharad, and Melanie Morten. 2019. "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia." *Journal of Political Economy*, 127(5): 2229–2268.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *The Review of Economic Studies*, 82(1): 1–44.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica*, 87(3): 741–835.

Carvalho, Vasco M, Makoto Nirei, Yukiko U Saito, and Alireza Tahbaz-Salehi. 2021. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake*." *The Quarterly Journal of Economics*, 136(2): 1255–1321.

Castro-Vincenzi, Juanma, Eugenia Menaguale, Eduardo Morales, and Alejandro Sabal. 2024. “Market Entry and Plant Location in Multi-Product Firms.” Working Paper.

Cavallo, Alberto, Gita Gopinath, Brent Neiman, and Jenny Tang. 2021. “Tariff Pass-Through at the Border and at the Store: Evidence from US Trade Policy.” *American Economic Review: Insights*, 3(1): 19–34.

Chalendard, Cyril, Ana M Fernandes, Gael Raballand, and Bob Rijkers. 2023. “Corruption in Customs*.” *The Quarterly Journal of Economics*, 138(1): 575–636.

Chaney, Thomas. 2008. “Distorted Gravity: The Intensive and Extensive Margins of International Trade.” *The American Economic Review*, 98(4): 1707–1721.

Chen, Hui, Yuhan Cheng, Yanchu Liu, and Ke Tang. 2026. “Teaching Economics to the Machines.” National Bureau of Economic Research Working Paper 34713.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. “Chapter 4 - Trade Theory with Numbers: Quantifying the Consequences of Globalization.” In *Handbook of International Economics*. Vol. 4 of *Handbook of International Economics*, , ed. Gita Gopinath, Elhanan Helpman and Kenneth Rogoff, 197–261. Elsevier.

Costinot, Arnaud, and Jonathan Vogel. 2010. “Matching and inequality in the world economy.” *Journal of Political Economy*, 118(4): 747–786.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams. 2019. “The More We Die, The More We Sell? A Simple Test of the Home-Market Effect*.” *The Quarterly Journal of Economics*, 134(2): 843–894.

Dai, Yue, Minhao Chen, and Zhengyu Zuo. 2023. “Neighbors in Space: Satellite Imagery and Chinese B-share Discount.” *China Economic Review*, 82: 102063.

Davis, Donald R, and James Harrigan. 2011. “Good jobs, bad jobs, and trade liberalization.” *Journal of international Economics*, 84(1): 26–36.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2007. “Unbalanced Trade.” *NBER working paper 13035*.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik. 2016. “Prices, Markups, and Trade Reform.” *Econometrica*, 84(2): 445–510.

de Roux, Nicolás, Marcela Eslava, Santiago Franco, and Eric A. Verhoogen. 2021. “Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments.”

Dhyne, Emmanuel, Ayumu Ken Kikkawa, Magne Mogstad, and Felix Tintelnot. 2021. “Trade and Domestic Production Networks.” *The Review of Economic Studies*, 88(2): 643–668.

Dickstein, Michael J., and Eduardo Morales. 2018. “What Do Exporters Know?” *Quarterly Journal of Economics*, 133(4).

Dingel, Jonathan I, and Felix Tintelnot. 2023. “Spatial Economics for Granular Settings.”

Disdier, Anne-Célia, and Keith Head. 2008. “The Puzzling Persistence of the Distance Effect on Bilateral Trade.” *The Review of Economics and Statistics*, 90(1): 37–48.

Dix-Carneiro, Rafael, and Brian K Kovak. 2019. “Margins of labor market adjustment to trade.” *Journal of International Economics*, 117: 125–142.

Dix-Carneiro, Rafael, João Paulo Pessoa, Ricardo Reyes-Heroles, and Sharon Traiberman. 2023. "Globalization, Trade Imbalances, and Labor Market Adjustment*." *The Quarterly Journal of Economics*, 138(2): 1109–1171.

Dix-Carneiro, Rafael, Pinelopi K Goldberg, Costas Meghir, and Gabriel Ulyssea. 2021. "Trade and Informality in the Presence of Labor Market Frictions and Regulations."

Djankov, Simeon, Caroline Freund, and Cong S Pham. 2010. "Trading on Time." *The Review of Economics and Statistics*, 92(1): 166–173.

Donaldson, Dave. 2008. "Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure."

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, 108(4-5): 899–934.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography and Trade." *Econometrica*, 70(5): 1741–1779.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2011. "An Anatomy of International Trade: Evidence from French Firms." *Econometrica*, 79(5): 1453–1498.

Faber, Benjamin, and Cecile Gaubert. 2019. "Tourism and Economic Development: Evidence from Mexico's Coastline." *American Economic Review*, 109: 2245–93.

Fajgelbaum, Pablo, and Stephen J. Redding. 2022. "Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914." *Journal of Political Economy*, 130(5): 1249–1318.

Fajgelbaum, Pablo, Cecile Gaubert, Nicole Gorton, Eduardo Morales, and Edouard Schaal. 2023. "Political Preferences and Transport Infrastructure: Evidence from California's High-Speed Rail." National Bureau of Economic Research w31438, Cambridge, MA.

Fajgelbaum, Pablo D, and Cecile Gaubert. 2020. "Optimal spatial policies, geography, and sorting." *The Quarterly Journal of Economics*, 135(2): 959–1036.

Fajgelbaum, Pablo D, Pinelopi K Goldberg, Patrick J Kennedy, and Amit K Khandelwal. 2020. "The Return to Protectionism*." *The Quarterly Journal of Economics*, 135(1): 1–55.

Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman. 2015. "A Linder Hypothesis for Foreign Direct Investment." *The Review of Economic Studies*, 82(1): 83–121.

Fernández-Villaverde, Jesús, Yiliang Li, Le Xu, and Francesco Zanetti. 2025. "Charting the Uncharted: The (Un)Intended Consequences of Oil Sanctions and Dark Shipping." University of Pennsylvania.

Fisman, Raymond, and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence from Missing Imports in China." *Journal of Political Economy*, 112(2): 471–500.

Flaaen, Aaron B, Ali Hortaçsu, Felix Tintelnot, Nicolás Urdaneta, and Daniel Xu. 2025*a*. "Who Pays for Tariffs Along the Supply Chain? Evidence from European Wine Tariffs." National Bureau of Economic Research.

Flaaen, Aaron B., Fariha Kamal, Eunhee Lee, and Kei-Mu Yi. 2025*b*. "An Anatomy of U.S. Establishments' Trade Linkages in Global Value Chains." Issue: 33724 Series: Working Paper Series Type: Working Paper _eprint: 33724.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth." *The American Economic Review*, 89(3): 379–399.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi. 2023. "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade." *The Review of Economic Studies*, 90(1): 331–375.

Ganapati, Sharat, and Colin J Hottman. 2026. "Did Foreigners Pay America's Tariffs? Quantity Discounts, Scale Economies and Incomplete Pass-Through." National Bureau of Economic Research Working Paper 34901.

Garcia-Marin, Alvaro, and Nico Voigtländer. 2019. "Exporting and Plant-Level Efficiency Gains: It's in the Measure." *Journal of Political Economy*, 127(4): 1777–1825.

Goldberg, Penny, and Giovanni Maggi. 1999. "Protection for Sale: An Empirical Investigation." *American Economic Review*, 89(5): 1135–1155.

Goldberg, Pinelopi K., Amit Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *Quarterly Journal of Economics*, 125(4): 1727–1767.

Goldberg, Pinelopi K, and Nina Pavcnik. 2016. "The Effects of Trade Policy." In *Handbook of Commercial Policy*. Vol. 1, 161–206. Elsevier.

Goldsmith-Pinkham, Paul. 2024. "Tracking the Credibility Revolution across Fields." Yale School of Management.

Gopinath, Gita, and Brent Neiman. 2014. "Trade Adjustment and Productivity in Large Crises." *American Economic Review*, 104(3): 793–831.

Grossman, Gene M., and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review*, 84(4): 833–850.

Handley, Kyle, and Nuno Limão. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *American Economic Review*, 107(9): 2731–2783.

Hanson, Gordon H., and Chong Xiang. 2004. "The Home-Market Effect and Bilateral Trade Patterns." *American Economic Review*, 94(4): 1108–1129.

Head, Keith, and Thierry Mayer. 2013. "Gravity Equations: Toolkit, Cookbook, Workhorse." In *Handbook of International Economics*. Vol. 4, , ed. Gita Gopinath, Elhanan Helpman and Kenneth Rogoff. New York: Elsevier.

Helpman, Elhanan. 1998. "The Structure of Foreign Trade."

Helpman, Elhanan, and Oleg Itskhoki. 2010. "Labor Market Rigidities, Trade and Unemployment." *Review of Economic Studies*, 77(3): 1100–1137.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export Versus FDI with Heterogeneous Firms." *The American Economic Review*, 94(1): 300–316.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *Quarterly Journal of Economics*, 2(5): 441–487.

Helpman, Elhanan, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J. Redding. 2017. "Trade and Inequality: From Theory to Estimation." *The Review of Economic Studies*, 84(1): 357–405.

Helpman, E., O. Itskhoki, and S. Redding. 2010. "Inequality and Unemployment in a Global Economy." *Econometrica*, 78(4): 1239–1283.

Hertel, T. W. 1997. *Global Trade Analysis: Modeling and Applications.* Cambridge:Cambridge University Press.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity *." *The Quarterly Journal of Economics*, 131(3): 1291–1364.

Hummels, David L., and Georg Schaur. 2013. "Time as a Trade Barrier." *American Economic Review*, 103(7): 2935–2959.

Huo, Zhen, Andrei A. Levchenko, and Nitya Pandalai-Nayar. 2025. "International Comovement in the Global Production Network." *The Review of Economic Studies*, 92(1): 365–403.

Javorcik, Beata Smarzynska. 2004. "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages." *American Economic Review*, 94(3): 605–627.

Johnson, Robert C., and Guillermo Noguera. 2012. "Accounting for Intermediates: Production Sharing and Trade in Value Added." *Journal of International Economics*, 86(2): 224–236.

Juhász, Réka. 2018. "Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade." *American Economic Review*, 108(11): 3339–76.

Kashiwagi, Yuzuka, Yasuyuki Todo, and Petr Matous. 2021. "Propagation of Economic Shocks through Global Supply Chains—Evidence from Hurricane Sandy." *Review of International Economics*, 29(5): 1186–1220.

Kehoe, Timothy J. 2005. "An Evaluation of the Performance of Applied General Equilibrium Models on the Impact of NAFTA." In *Frontiers in Applied General Equilibrium Modeling.* , ed. Timothy J. Kehoe, T. N. Srinivasan and John Whalley, 341–377. Cambridge University Press.

Kehoe, Timothy J., Clemente Polo, and Ferran Sancho. 1995. "An Evaluation of the Performance of Applied General Equilibrium Models of the Impact of NAFTA." *Moneda y Credito*, 201: 13–64.

Kehoe, Timothy J., Pau S. Pujolas, and Jack Rossbach. 2017. "How Important is the New Goods Margin in International Trade?" *Journal of Political Economy*, 125(2): 358–392.

Kikuchi, Shinnosuke. 2025. "Does Skill Abundance Still Matter? The Evolution of Comparative Advantage in the 21st Century." MIT.

Kleinman, Benny, Ernest Liu, and Stephen J. Redding. 2024. "International Friends and Enemies." *American Economic Journal: Macroeconomics*, 16(4): 350–385.

Koopman, Robert, Zhi Wang, and Shang-Jin Wei. 2014. "Tracing Value-Added and Double Counting in Gross Exports." *American Economic Review*, 104(2): 459–494.

Kotia, Ananya. 2025. "Family Matters: Globalization, Management, and Firm Productivity." London School of Economics.

Kovak, Brian K. 2013. "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" *American Economic Review*, 103(5): 1960–1976.

Kugler, Maurice, and Eric Verhoogen. 2012. "Prices, Plant Size and Product Quality." *Review of Economic Studies*, 79(1): 307–339.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh. 2023. "The welfare effects of encouraging rural–urban migration." *Econometrica*, 91(3): 803–837.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review*, 76(4): 604–620.

Lane, Nathan. 2025. "Manufacturing revolutions: Industrial policy and industrialization in South Korea." *The Quarterly Journal of Economics*, qjaf025.

Leontief, Wassily. 1953. "Domestic production and foreign trade; the American capital position re-examined." *Proceedings of the American philosophical Society*, 332–349.

Liu, Ernest. 2019. "Industrial Policies in Production Networks."

Mayer, Thierry, Marc J Melitz, and Gianmarco IP Ottaviano. 2021. "Product mix and firm productivity responses to trade competition." *Review of Economics and Statistics*, 103(5): 874–891.

McCallum, John. 1995. "National Borders Matter: Canada-U.S. Regional Trade Patterns." *The American Economic Review*, 85(3): 615–623.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2008. "Market Size, Trade, and Productivity." *The Review of Economic Studies*, 75(1): 295–316.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review*, 108(12): 3855–3890.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives*, 31(2): 87–106.

Nunn, Nathan. 2007. "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade*." *The Quarterly Journal of Economics*, 122(2): 569–600.

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *The Review of Economic Studies*, 69(1): 245–276.

Ramondo, Natalia, and Andres Rodríguez-Clare. 2013. "Trade, Multinational Production, and the Gains from Openness." *Journal of Political Economy*, 121(2): 273–322.

Ramondo, Natalia, Andres Rodríguez-Clare, and Felix Tintelnot. 2015. "Multinationals and the Gains from International Trade." *Review of Economic Studies*. NBER Working Paper 21321.

Ramondo, Natalia, Veronica Rappoport, and Kim J. Ruhl. 2016. "Intrafirm Trade and Vertical Fragmentation in U.S. Multinational Corporations." *Journal of International Economics*, 98: 51–59.

Rauch, James E., and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade." *The Review of Economics and Statistics*, 84(1): 116–130.

Redding, Stephen J. 2016. "Goods trade, factor mobility and welfare." *Journal of International Economics*, 101: 148–167.

Redding, Stephen J., and Daniel M. Sturm. 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review*, 98(5): 1766–1797.

Roberts, Mark J., and James R. Tybout. 1997. "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *The American Economic Review*, 87(4): 545–564.

Sampson, Thomas. 2014. "Selection into Trade and Wage Inequality." *American Economic Journal: Microeconomics.*

Schott, Peter K. 2004. "Across-Product versus Within-Product Specialization in International Trade." *Quarterly Journal of Economics*, 119(2): 647–678.

Sequeira, Sandra. 2016. "Corruption, Trade Costs, and Gains from Tariff Liberalization: Evidence from Southern Africa." *American Economic Review*, 106(10): 3029–3063.

Shapiro, Joseph S. 2016. "Trade Costs, CO2, and the Environment." *American Economic Journal: Economic Policy*, 8(4): 220–254.

Shoven, John B., and John Whalley. 1984. "Applied General-Equilibrium Models of Taxation and International Trade: An Introduction and Survey." *Journal of Economic Literature*, 22(3): 1007–1051.

Simonovska, Ina, and Michael Waugh. 2011. "The Elasticity of Trade: Estimates and Evidence." *mimeo.*

Sotelo, Sebastian. 2020. "Domestic Trade Frictions and Agriculture." *Journal of Political Economy*, 128(7): 2690–2738.

Startz, Meredith. 2021. "The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade." Dartmouth Working Paper.

Teti, Feodora. 2024. "Missing Tariffs."

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, migration, and productivity: A quantitative analysis of China." *American Economic Review*, 109(5): 1843–1872.

Topalova, Petia. 2007. "Trade Liberalization, Poverty and Inequality: Evidence from Indian Districts." In *Globalization and Poverty.* , ed. Ann Harrison, 291–332. Chicago IL:University of Chicago Press.

Topalova, Petia, and Amit Khandelwal. 2011. "Trade liberalization and firm productivity: The case of India." *Review of economics and statistics*, 93(3): 995–1009.

Trefler, Daniel. 1995. "The Case of the Missing Trade and Other Mysteries." *American Economic Review*, 85(5): 1029–46.

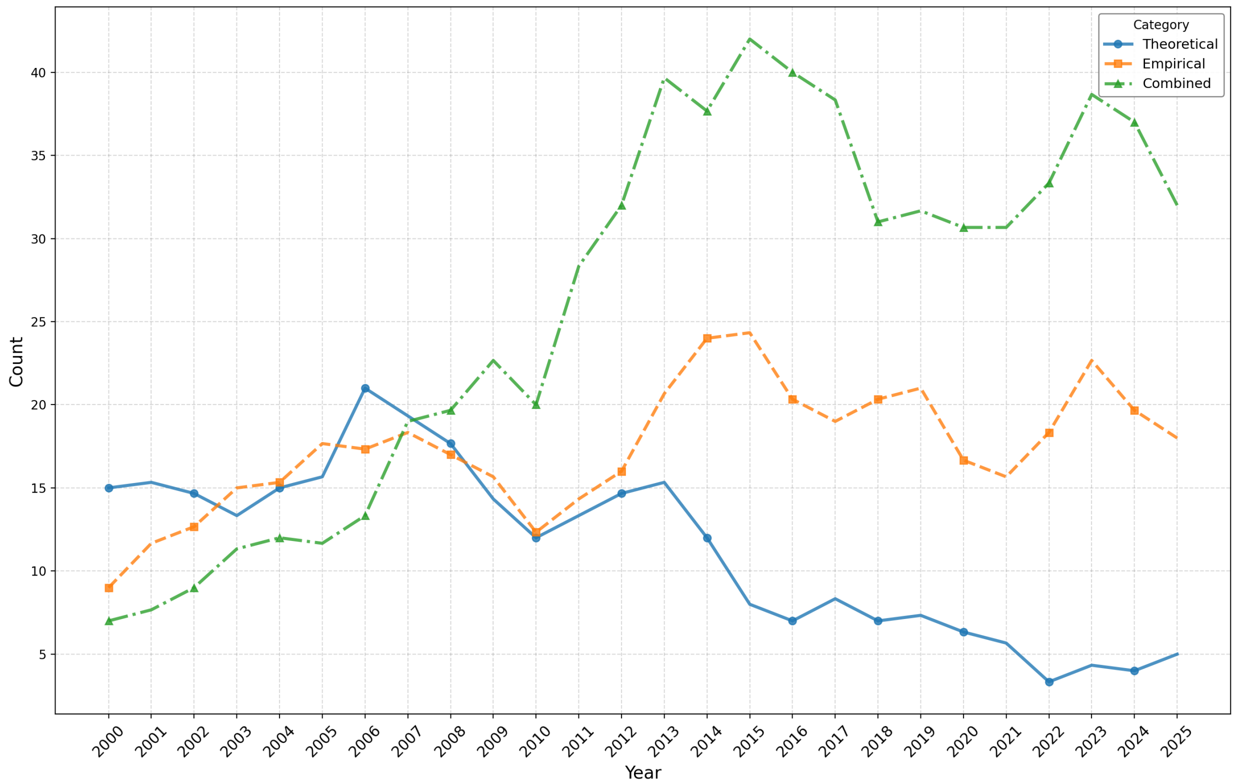
Tybout, James R., and M. Daniel Westbrook. 1995. "Trade Liberalization and the Dimensions of Efficiency Change in MexicanManufacturing Industries." *Journal of International Economics*, 39(1-2): 53–78.

Verhoogen, Eric. 2008. "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector." *Quarterly Journal of Economics*, 123(2): 489–530.

Walker, Michael W., Nachiket Shah, Edward Miguel, Dennis Egger, Felix Samy Soliman, and Tilman Graff. 2024. "Slack and Economic Development."

V. Appendix A: Additional Figures

Figure 7. : Papers by Methodology, 2000–2025: Paper Counts



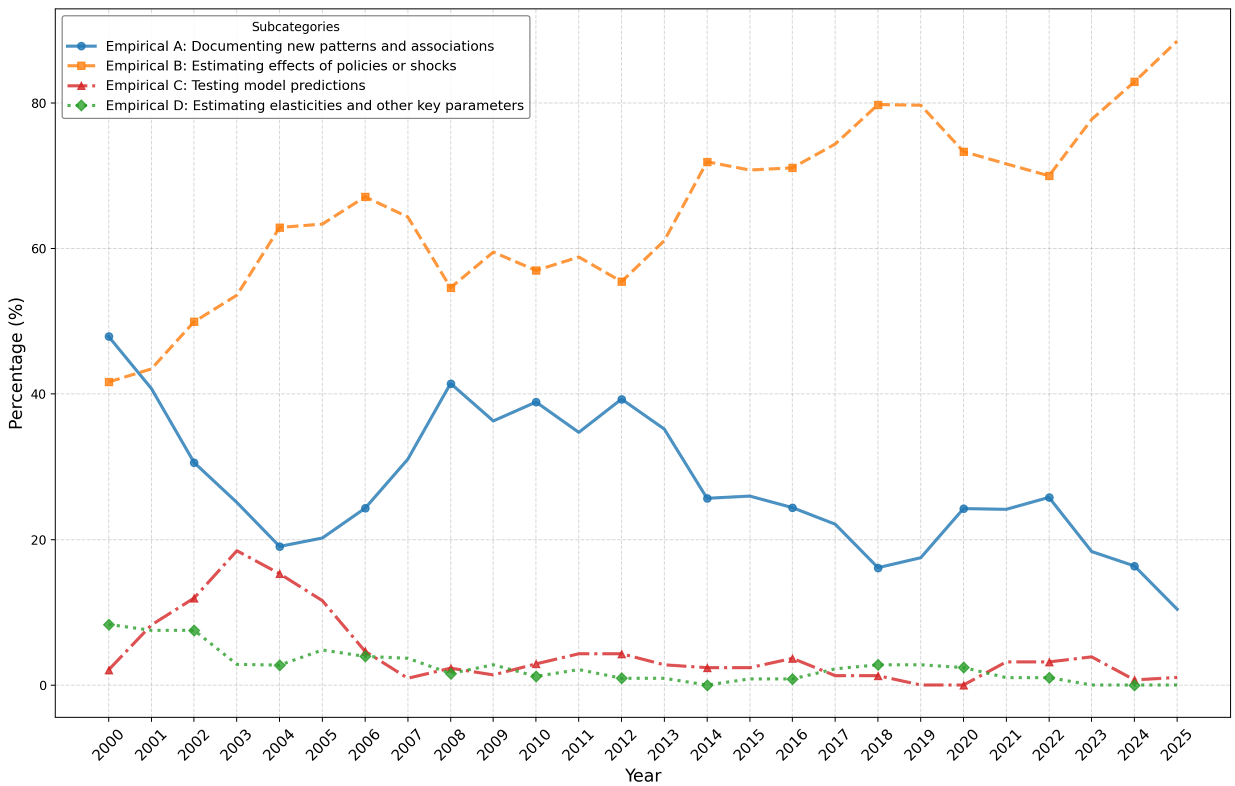
Notes: Figure 7 plots the count of papers in the field of international trade published in seven leading journals between 2000 and 2025 by broad category. Categorization done via LLMs through reading the first 2500 words of each paper. Lines are smoothed using a 3-year moving average.

Figure 8. : Composition of Papers by Journal Type

Category	Top 5 Journals	JIE
Theoretical	15.1%	25.5%
Empirical	25.4%	32.8%
Combined	59.4%	41.7%

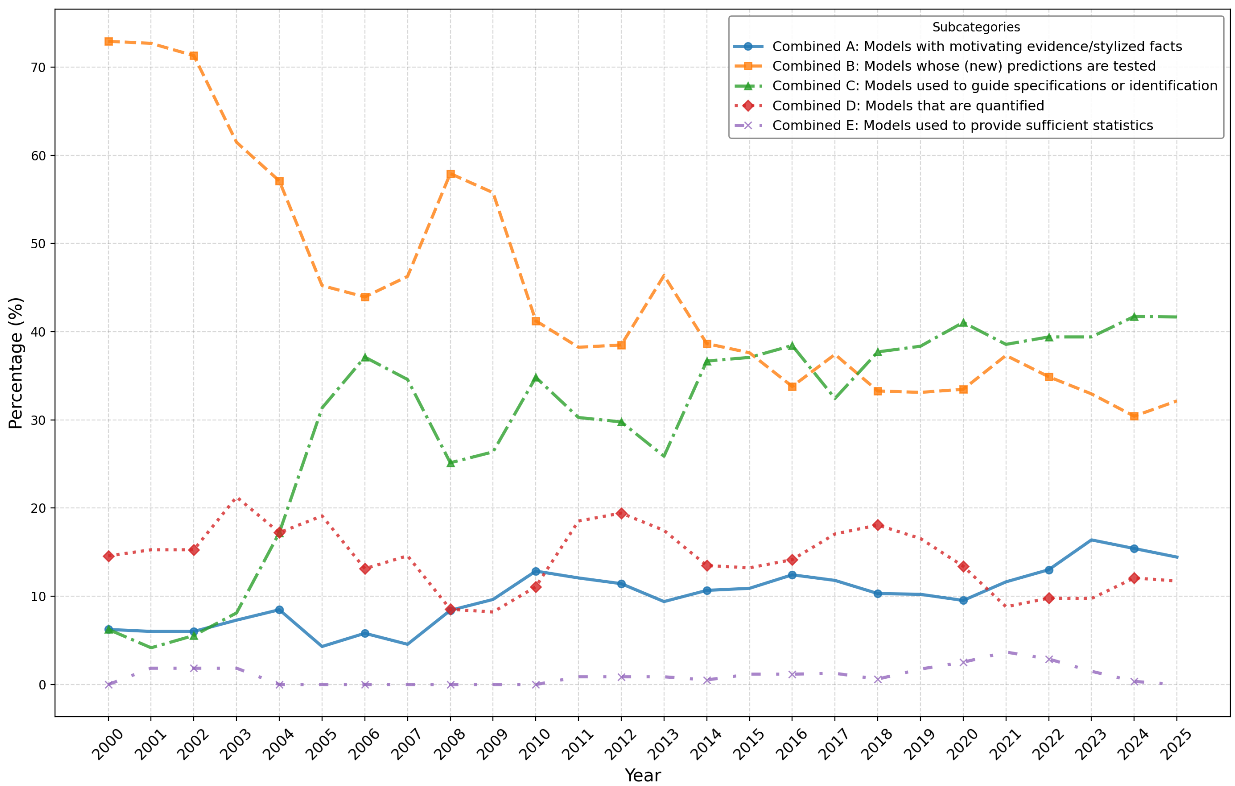
Notes: Figure 8 compares the share of papers in the field of international trade published in the Top 5 journals (AER, Econometrica, JPE, QJE, ReStud) and the Journal of International Economics (JIE) between 2000 and 2025. Categorization was done via LLMs through reading the first 2500 words of each paper.

Figure 9. : Evolution of Empirical Paper Subcategories 2000–2025: Fractional Shares



Notes: Figure 9 plots the share of empirical papers in the field of international trade published in seven leading journals between 2000 and 2025 by empirical subcategory. Categorization is done via LLMs through reading the first 2500 words of each paper. Lines are smoothed using a 3-year moving average. When a paper is assigned multiple subcategories, each one receives an equal fractional share such that the total contribution of each paper sums to one. Percentages sum to 100% in each year.

Figure 10. : Evolution of Combined Paper Subcategories 2000–2025: Fractional Shares



Notes: Figure 10 plots the share of combined theoretical and empirical papers in the field of international trade published in seven leading journals between 2000 and 2025 by combined subcategory. Categorization is done via LLMs through reading the first 2500 words of each paper. Lines are smoothed using a 3-year moving average. When a paper is assigned multiple subcategories, each one receives an equal fractional share such that the total contribution of each paper sums to one. Percentages sum to 100% in each year.

VI. Appendix B: LLM Prompts

Initial Classification:

You are an expert assistant specializing in analyzing economics research papers for an audience of academic economists. For each excerpt you are shown, classify it according to the following taxonomy. These categories are not mutually exclusive, so a paper can in principle fall in two categories. However, a paper should only be placed in multiple categories when the main contributions of the paper span multiple categories, not when in addition to their main contributions they also do something that falls into an additional category:

1) Determine the overall category of the paper (1, 2, or 3) based on the main contribution(s) as described by the authors — typically in the introduction (often begins with "In this paper...", "This paper investigates...", or "This paper attempts..."). The authors may also describe their contribution in any dedicated "Contribution" or "Related Literature" section that compares the paper's contribution to the existing literature.

2) If the paper falls under Category 2 or 3, assign at least one valid subcategory that reflects the nature of the main contribution(s). For Category 3 papers, subcategories (3.A–3.E) should describe how the empirical analysis is connected to the theoretical framework.

3) Provide a very concise explanation (two sentences max) for why this classification was made. Use direct quotes or close paraphrases from the paper to justify the decision. If this was a more challenging paper to classify because it did not neatly fit within category 1, 2 or 3, please add an additional sentence explaining that it was challenging and why.

Category 1 - Main contribution is theoretical

- The main contribution of the paper is theoretical, meaning that a new mathematical model of the economy is developed in the analysis.

- Papers in this category will frequently include in the introduction phrases like"we develop a model", "we develop a new model", "we develop a new theory" or an "we develop an extension of an existing theory", particularly in reference to the key contributions of the paper.

- These papers will devote most attention to describing and developing the theoretical framework, showing a number of mathematical equations that describe economic behavior, solving for relationships that hold in the model, presenting comparative statistics or , and discussing the new insights that emerge from these theoretical mechanisms. There will typically be combinations of lemma, propositions, or theorems in the paper.

- Good examples for papers where the main contribution is theoretical are "Trading tasks: A simple theory of offshoring" by Gene Grossman and Esteban Rossi-Hansberg, published in the American Economic Review in 2008, and "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity" by Marc Melitz, published in Econometrica in 2003.

- A key distinguishing feature is a theoretical paper will have no data or empirical work at all, or if it does it will be a few plots of associations early in the paper to motivate studying a subject or an assumption. However, any such analysis will be cursory at best and get little attention in the introduction.

- It is also possible that the primary contribution of a paper is econometric theory, i.e. a new estimator that has desirable statistical properties or a new way to interpret existing estimators. Good examples would be the "Bartik Instruments: What, When, Why, and How" paper by Paul Goldsmith-Pinkham, Isaac Sorkin and Henry Swift. These papers often talk about illustrating results through an application.

Category 2 - Main contribution is empirical

- The main contribution of the paper is empirical, meaning data analysis.
- If there is a model, it is not novel and may be described as a framework, and it plays only a very minor role in the paper and barely features in discussions of the contributions of the paper.
- These papers will not talk about quantifying a model, calibrating a model, or estimating a model.
- The paper can take various forms described below in categories 2.A to 2.D.
- Category 2.A - Main contribution is to use data to document new patterns and relationships
- The main contribution of the paper is empirical, meaning data analysis.
- The empirical analysis is mainly focused on documenting and describing new facts, patterns and relationships in the data.
- Papers in this category will frequently include in the introduction phrases like "new data", "new stylized facts", "new facts", "document new patterns", "document new relationship", "the analysis is mainly descriptive", particularly in reference to the key contributions of the paper.
- These papers will devote considerable attention to first describe the datasets used in the analysis, before proceeding to a number of new facts with figures and tables that summarize the relationships in the datasets. Rather than estimating effects, these papers focus on a more descriptive exercise that uses summary statistics and simple relationships that hold in the data. They will not devote meaningful attention to developing a theoretical model.
- Good examples for papers where documenting new patterns and relationships in the data are the main contribution are "Accounting for intermediates: Production sharing and trade in value added" by Robert Johnson and Guillermo Noguera, published in the Journal of International Economics in 2012, and "Dissecting Trade: Firms, Industries, and Export Destinations" by Jonathan Eaton, Samuel Kortum and Francis Kramarz, published in the American Economic Review in 2004.

Category 2.B - Main contribution is to use data to estimate the effects of policies or economic shocks

- The main contribution of the paper is empirical, meaning data analysis.
- The empirical analysis is mainly focused on estimating the causal effect of a policy reform or economic shock on economic outcomes, such as GDP, trade flows, wages, income and employment.
- Papers in this category will frequently include in the introduction phrases like "estimate the effect of", "identification strategy", "causal effect" and "natural experiment", particularly in reference to the key contributions of the paper.

- These papers will devote considerable attention to first describe the datasets used in the analysis, before proceeding to a regression analysis with tables and figures that summarize the point estimates and their statistical significance. They will not devote meaningful attention to developing a theoretical model.

- Good examples for papers where estimating the effects of policies or shocks using data are the main contribution are "The China syndrome: Local labor market effects of import competition in the United States" by David Autor, David Dorn and Gordon Hanson, published by the American Economic Review in 2013, "Distance, trade, and income-the 1967 to 1975 closing of the suez canal as a natural experiment" by James Feyrer, published in the Journal of Development Economics in 2021, and "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages" by Alfaro-Urena, Isabela Manelici and Jose Vasquez published in the Quarterly Journal of Economics in 2022.

Category 2.C - Main contribution is to use data to test predictions of an existing theoretical model

- The main contribution of the paper is empirical, meaning data analysis.

- The empirical analysis is mainly focused on testing the predictions of an existing theoretical model in the data.

- Papers in this category will frequently include in the introduction phrases like "in this paper I test the prediction that", "test of the model prediction", "test the model", "evaluate the comparative static" and "testing the theory", particularly in reference to the key contributions of the paper.

- These papers will devote considerable attention to first describe the predictions of the existing theory, describe the datasets used to test the predictions and then report the empirical specifications and the results in regression tables and figures.

- Key here is that the papers are testing an existing well-known model, not a new one that they have themselves come up with. Thus they will talk about testing a model or theory from another paper that they reference.

- Good examples for papers where testing the predictions of existing theories are the main contribution are "Across-Product Versus Within-Product Specialization in International Trade" by Peter Schott, published in the The Quarterly Journal of Economics in 2004, "Bones, Bombs, and Break Points: The Geography of Economic Activity" by Donald Davis and David Weinstein, published in the American Economic Review in 2002, and "Relationship-specificity, incomplete contracts, and the pattern of trade" by Nathan Nunn, published in the Quarterly Journal of Economics in 2007. "Protection for Sale: An Empirical Investigation" by Penny Goldberg and Giovanni Maggi, published in the American Economic Review in 1999 is another good example that seeks to test Grossman and Helpmans' Protection for Sale model (and also estimates the weight placed on welfare versus lobbying contributions that falls in category 2.D).

Category 2.D - Main contribution is to use data to estimate an elasticity or another parameter that has been shown to be important in an existing theoretical model

- The main contribution of the paper is empirical, meaning data analysis.

- The empirical analysis is mainly focused on estimating a parameter of interest from existing theoretical models, including elasticities that describe how outcomes change as a function of another variable.

- Papers in this category will frequently include in the introduction phrases like “estimate this parameter” and “provide an estimate of this elasticity”, particularly in reference to the key contributions of the paper.

- These papers will devote considerable attention to first describe the datasets used in the analysis, before proceeding to the empirical analysis that provides a number of estimates of the key parameters or elasticities of interest. Often they will lay out the existing model in order to motivate their specification or highlight the parameter they are estimating.

- Good examples for papers where estimating important parameters or elasticities from existing theoretical models are the main contribution are “The long and short (run) of trade elasticities” by Christoph Boehm, Andrei Levchenko and Nittya Pandalai-Nayar, published in the American Economic Review in 2023, and “The Puzzling Persistence of the Distance Effect on Bilateral Trade’ by Anne-Célia Disdier and Keith Head, published in the Review of Economics and Statistics in 2008. “Protection for Sale: An Empirical Investigation” by Penny Goldberg and Giovanni Maggi, published in the American Economic Review in 1999 is another good example that seeks to estimate the weight placed on welfare versus lobbying contributions (but also tests the model).

Category 3 - Main contribution combines theory with data

- The main contributions of the paper include both theoretical and empirical contributions, meaning developing a theoretical model and combining this with data analysis.

- The paper can take various forms described below in categories 3.A to 3.E.

- Category 3.A - Main contribution combines theory with data to show evidence that motivates the development of a new theoretical model

- The paper combines theoretical and empirical contributions, meaning developing a theoretical model and combining this with data analysis.

- The data analysis in this category of papers is mainly used to present motivating evidence that informs the structure and assumptions of the theoretical model.

- Papers in this category will frequently include in the introduction phrases like “present stylized facts that inform the theoretical model” and “present motivating evidence to inform the theory”, particularly in reference to the key contributions of the paper.

- These papers will devote considerable attention to first presenting a number of new stylized facts that are estimated using datasets before developing a new theoretical model motivated by those facts or using assumptions based on those facts.

- Good examples for papers where using data for motivating evidence to develop a theoretical model are the main contribution is “A Linder Hypothesis for Foreign Direct Investment” by Pablo Fajgelbaum, Gene Grossman and Elhanan Helpman, published in the Review of Economic Studies in 2015. In that paper, the evidence presented in Figure 1 is used to motivate assumptions and features of the new theoretical model. There are also many papers that present stylized facts that motivate the model or model assumptions before going on to perform other steps discussed in Categories 3.B, 3.C and 3.D, for example “Volatility and the Gains From Trade” by Treb Allen and David Atkin published in Econometrica in 2022.

Category 3.B - Main contribution combines theory with data to first write down a new theoretical model and then test some of the key predictions of that theory using data

- The paper combines theoretical and empirical contributions, meaning developing a theoretical model and combining this with data analysis.
- The data analysis in this category of papers is mainly used to test the predictions of the theoretical model.
- Papers in this category will frequently include in the introduction phrases like "test the predictions of our model", "explore the comparative statics of the model" and "testing the theory", particularly in reference to the key contributions of the paper.
- In contrast to Category 2.C, here the model being tested is a new model introduced in the paper.
- These papers will devote most attention to developing the theoretical model, derive predictions on observable outcomes, then discuss how they test those predictions using datasets and some form of statistical analysis.
- Good examples for papers where using data to test the predictions of a new theoretical model are one of the main contributions are "Multiproduct firms and trade liberalization" by Andrew Bernard, Stephen Redding and Peter Schott, published in the Quarterly Journal of Economics in 2011, and "The home-market effect and bilateral trade patterns" by Gordon Hanson and Chong Xiang, published in the American Economic Review in 2004.

Category 3.C - Main contribution combines theory with data to develop the theory in order to guide specifications of the empirical analysis

- The paper combines theoretical and empirical contributions, meaning developing a theoretical model and combining this with data analysis.
- The data analysis in this category of papers is mainly used to derive the specifications for the empirical analysis, such as regression specifications, or to discuss threats to identification.
- Papers in this category will frequently include in the introduction phrases like "the model guides the empirical analysis" and "derive the specification from the model", "helps clarify our identification strategy" particularly in reference to the key contributions of the paper.
- These papers will devote considerable attention to presenting the theoretical framework, then derive the main specifications that will be used in the empirical analysis using datasets and regression analysis.
- Good examples for papers where using the theoretical model to guide the empirical analysis are the main contribution are "Railroads of the Raj: Estimating the impact of transportation infrastructure" by Dave Donaldson, published in the American Economic Review in 2018,"Regional effects of trade reform: What is the correct measure of liberalization?" by Brian Kovak, published in the American Economic Review in 2013, and "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas" by Arnaud Costinot, Dave Donaldson and Ivana Komunjer published in the Review of Economic Studies in 2012.

Category 3.D - Main contribution combines theory with data to develop the theory and use the data to calibrate and quantify the model and to perform counterfactuals

- The paper combines theoretical and empirical contributions, meaning developing a theoretical model and combining this with data analysis.
- The data analysis in this category of papers is mainly used to quantify the theoretical model. Calibration or estimation may be used to obtain the model parameters.

- Papers in this category will frequently include in the introduction phrases like "develop and quantify a model", "calibrate the model" "quantitative assessment", "perform counterfactuals", particularly in reference to the key contributions of the paper.
- These papers will devote considerable attention to develop the theoretical model, introduce the data they use to calibrate the model or to estimate the key parameters, and then at the end of the paper the papers will quantify the model to assess the importance of various mechanisms, and/or perform counterfactuals using the estimated or calibrated model.
- Good examples for papers where using data to calibrate the theoretical model are the main contribution are "Technology, Geography and Trade" by Jonathan Eaton and Samuel Kortum published in Econometrica in 2002, "An Anatomy of International Trade: Evidence From French Firms" by Jonathan Eaton, Samuel Kortum and Francis Kramarz, published in Econometrica in 2011, "Estimates of the Trade and Welfare Effects of NAFTA" by Lorenzo Caliendo and Fernando Parro, published in the Review of Economic Studies in 2015, "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914" by Pablo Fajgelbaum and Stephen Redding published in the Journal of Political Economy in 2022.

Category 3.E - Main contribution combines theory with data by first deriving from the theory a small set of sufficient statistics (those are the key parameters that one needs to measure in the data to evaluate the effects of a policy) that hold in a broad set of models, and then using data to estimate those statistics as credibly as possible to answer a question.

- The paper combines theoretical and empirical contributions, meaning developing a theoretical model and combining this with data analysis.
- The paper writes down a set of models or a very general theoretical structure and derives sufficient statistics that allow answers to key policy questions without needing to commit to a single model.
- Papers in this category will typically refer to "sufficient statistics" in their introduction, typically citing the review by Chetty (2009) titled "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods", particularly in reference to the key contributions of the paper.
- These papers will typically devote a fair amount of space to showing how many models or a broad class of models produce answers to policy questions that depend on a few observable moments such as elasticities or shares. These moments may be endogenous objects in the theory rather than primitives. Then they will estimate those key objects before addressing the policy question for which those moments provide sufficient statistics.
- Good examples in trade include "New Trade Models, Same Old Gains?" by Arkolakis, Costinot, and Rodríguez-Clare (2012) which shows that the share of expenditure on domestic consumption and the trade elasticity is sufficient to calculate the gains from trade in a range of models, Adao, Arkolakis and Esposito (2024) in the "General Equilibrium Effects in Space", Kleinman, Liu and Redding (2024) in "International Friends and Enemies", and Huo, Levchenko and Pandalai-Nayar (2025) "International Comovement in the Global Production Network".

Based on these descriptions, classify the paper into one of three main categories — and, if the paper is classified as Category 2 or Category 3, into one or more valid subcategories.

You must respond only with a strict JSON object in the following format:

{

"category": "1" | "2" | "3",

“subcategory”: [“2.A”, “2.B”], // REQUIRED if category is 2 or 3; must match category
“justification”: “The authors state, ‘We estimate the causal impact of education on earnings...’ which supports an empirical classification.”
}
Rules:
- If the paper is classified as Category 1, set “subcategory” to ‘null’.
- If the paper is classified as Category 2 or Category 3, you must assign at least one valid subcategory.
- Assign all subcategories that are supported by the content. Subcategories are not mutually exclusive.
- Each subcategory must start with the same number as the main category (e.g., 3.A–3.E for Category 3).
- Only assign Category 3 if the theoretical contribution is substantial and clearly integrated with the empirical work. A brief or minor model is not sufficient.
- Do **not** leave the subcategory field empty or null when the category is 2 or 3.
- The justification must be clear and specific. Reference the content of the excerpt.

Adjudication Prompt:

You are acting as a two-agent system specializing in analyzing economics research papers, designed to provide an authoritative classification for research papers where models have disagreed.

First agent: Your task is to read the previous justifications written by both Opus and Gemini, and the extracted sample of 2,500 words from each paper in extracted_texts_full.json, and look for specific methodological evidence. Is the paper’s main contribution theoretical, empirical or a combination of theory and empirics? Do not classify the paper yet.

Second agent: Using the evidence extracted by Agent 1 and the following taxonomy, provide the final classification (category and subcategory). (Same taxonomy as initial prompt)

Based on these descriptions, classify the paper into one of three main categories — and, if the paper is classified as Category 2 or Category 3, into one or more valid subcategories.

You must respond only with a strict JSON object in the following format:
{
“final_category”: “1, 2, or 3”,
“final_subcategories”: [“list”, “of”, “subcats”],
}

Unit of Analysis Prompt:

You are an expert assistant specializing in analyzing economics research papers for an audience of academic economists. Your task is to identify the PRIMARY unit of analysis in a research paper’s main empirical or quantitative section. The “unit of analysis” is the specific level of observation at which the main regressions are run.

You must choose only from the list below. Use these exact strings, and only include data types that are central to the paper’s empirical contribution. Do not list every possible dataset mentioned in the paper in passing—only those that are used in the main analysis.

Valid data types (choose ONE):

- “country”: Cross-country analysis (e.g., each observation is a nation-year).
- “region within a country”: Sub-national units like states, provinces, cities, or counties.
- “sector within a country”: Industry-level data (e.g., ISIC or NAICS codes) within one or more nations.

- “firm or establishment”: Micro-data on individual companies, plants, or businesses.
- “household or individual”: Consumer-level, worker-level, or survey-based individual data.
- “product or commodity”: High-resolution trade data (e.g., HS 6-digit codes) where the product is the observation.
- “country pair”: Bilateral trade flows or gravity-style analysis (e.g., exports from Country A to Country B).
- “not applicable”: Use ONLY if the paper is purely theoretical with no data.

Rules:

1) Identify the level of the MAIN regression or calibration. Ignore secondary robustness checks or preliminary data descriptions.
2) If the data is multi-level (e.g., firm-level data aggregated to the sector level for the main result), choose the level of the final estimation.
3) A paper studying international trade flows between specific origin and destination pairs should be classified as country pair.

You must respond only with a strict JSON object in the following format:
json
{
unit_of_analysis: <one of the valid units listed above>,
}

Methodology and Estimator Prompt:

You are an expert assistant specializing in analyzing economics research papers for an audience of academic economists. Your goal is to identify the identification strategy and the estimators used in the main empirical analysis.

You must return a JSON object with the following structure:
{
strategy: [strategy_1, strategy_2],
“estimator”: [estimator_1, estimator_2],
justification: Short explanation, based on evidence from the paper, of why the identification strategies and estimators listed are appropriate.
}

If the paper is purely theoretical and does not contain any empirical estimation or data analysis, return:
{
strategy: [],
justification: Short explanation, based on evidence from the paper, of why the paper is purely theoretical
}

You must choose only from the list below. Use these exact strings, and only include strategies that are central to the paper’s main empirical contribution. Only list methods that play a central role in the empirical analysis.

Core identification designs:

- difference-in-differences – The paper compares changes over time between treated and control groups to identify causal effects. Papers that use regression analysis based on difference in differences often mention phrases such as "to identify the causal effect we use an empirical strategy based on difference in differences.", sometimes abbreviated to DiD. Often there will be a binary treatment indicator interacted with a post indicator used as the coefficient of interest. Also include "triple difference" designs here (aka difference-in-difference-in-difference).

- event study – The paper estimates treatment effects across time relative to a discrete event, often visualized with leads and lags of the treatment effects. The paper will often mention phrases such as "using an event study design". Discussion that identification comes from variation in timing relative to a single event date.

- "instrumental variables designs – The paper uses instrumental variables to uncover causal relationships when the treatment is endogenous. This includes the use of shift-share or Bartik-style instrumental variables to identify causal effects.

- regression discontinuity design – The paper identifies causal effects by exploiting a cutoff in an assignment variable. There may be explicit mention of a running variable and a cutoff that determines treatment as well as potentially language describing bandwidth choice, or "sharp vs fuzzy RD."

- randomized control trial – The paper uses an actual experiment to randomly assign treatment and control groups to estimate causal effects. There is typically a clear statement of random assignment procedure and phrases like "randomized," "random assignment," or "randomization protocol.

- "matching estimators" – This covers a range of methods including matching estimators using propensity scores, reweighting approaches using inverse propensity weighting and synthetic controls. The core identification strategy is to use observables to help select a control group that has balanced covariates with the treatment group in the absence of experimental or quasi experimental assignment. For example, matching estimators pick controls based on similarity with the treated units, synthetic control methods a paper constructs a weighted combination of control units to act as a counterfactual for the treated unit. Papers that use synthetic controls may reference Abadie et al. (2010). Papers that use matching estimators frequently include terms such as propensity score matching. Papers that use reweighting frequently include terms like "inverse probability weights". Do not include papers here that simply control for observable confounds.

- "no identification design" – This covers papers that do not claim to causally identify impacts or parameters. This also covers papers that are vague regarding casualty and identification or simply rely on controls for observables or fixed effects to try to reduce bias. These papers will typically not talk about identification, threats to identification, confounds or causality. If there is any discussion it is extremely brief and mainly talks about controls and fixed effects as the solution.

Core estimators:

- Ordinary Least Squares (OLS) – The paper uses ordinary least squares or linear regression. Papers that use OLS will often include phrases like "we estimate an OLS regression..". Typically fixed effects or basic panel structures will use OLS. The paper does not specify that it is using one of the other methods described below that also admit fixed effects. "Two Stage Least Squares (2SLS)" should also be grouped with OLS. These papers use an instrumental variable design implemented via two stage least squares. Often the paper will talk about using an instrument or an IV strategy and not additionally discuss one of the other estimators below (as you can also use instrumental variables with the approaches below such as GMM).

- "generalized method of moments (GMM)" – The paper uses moment conditions to estimate parameters, often in the presence of endogeneity. Papers that use GMM will often include phrases like \two-step GMM\, or \we estimate using the generalized method of moments\.,
- "poisson pseudo-maximum likelihood" – The paper uses Poisson regression methods, often for count data or in gravity models of trade. Papers that use PPML will often include phrases like \Poisson pseudo-maximum likelihood\, \PPML\, \count data regression\, or \gravity equation estimated with Poisson\.,
- "logit and probit estimators" – The paper uses logistic or probit regression to model binary outcomes. Papers that use these estimators will often include phrases like \logit regression\, \probit model\, \binary dependent variable\, \logistic regression\,
- "multinomial and ordered choice estimators" – The paper uses multinomial logit, multinomial probit, ordered logit, or ordered probit models to analyze discrete choices with multiple categories. Papers using these estimators will often include phrases like \multinomial logit\, \ordered probit\, \discrete choice model\, \multiple categories\, or \rank-ordered outcomes\.,
- "quantile regression" – The paper estimates effects at different points of the conditional distribution rather than just the mean. Papers that use quantile regression will often include phrases like \quantile regression\, \median regression\, \conditional quantiles\.,
- machine learning methods – The paper uses machine learning techniques for prediction or estimation. The paper might include the words or phrases: "random forest", "neural network", "double debiased machine learning", "causal forests" "we use machine learning to predict…",
- "duration and hazard models" – The paper uses duration models where the object of interest is the time until failure, or exit or closure. The paper often uses phrases sychs as "hazard rate", "baseline hazard", "survival function".
- "bayesian estimators" – including Bayesian MCMC-based estimators, Bayesian minimum distance approaches
- "nonparametric and semi-parametric estimators" – including Kernel, Series, Local polynomial
- "maximum likelihood estimation (MLE)" – The paper estimates parameters by maximizing a likelihood function that represents the probability of observing the data given the parameters. Papers using MLE will often include phrases like \maximum likelihood\, \likelihood function\, \we maximize the likelihood\, \MLE\, or \log-likelihood\.,
- "simulated method of moments (SMM)" – The paper estimates model parameters by matching moments from simulated data to moments from actual data. Papers using SMM will often include phrases like \simulated method of moments\, \match simulated moments to data moments\.,
- "indirect inference" – The paper estimates structural parameters by matching statistics or parameter estimates from an auxiliary model fitted to simulated and actual data. Papers using indirect inference will often include phrases like \indirect inference\, \auxiliary model\, \binding function\, or \match auxiliary parameters\.,

- “minimum distance estimators” – The paper estimates parameters by minimizing a weighted distance between sample moments and their theoretical counterparts. Papers using minimum distance methods will often include phrases like \minimum distance\, \minimum chi-squared\, \optimal weighting matrix\.,
- “calibration-based estimators (when parameters are disciplined by matching moments)” – The paper sets model parameters to match specific empirical moments or targets from the data. Papers using calibration will often include phrases like \calibrate the model to match\, \set parameters to target\, \calibration exercise\, or \externally calibrated\.,
- “regularization methods (LASSO, Post-LASSO)” – The paper uses penalized regression methods that shrink or select variables to prevent overfitting. Papers using regularization methods will often include phrases like \LASSO\, \ridge regression\, \elastic net\, \penalized regression\, \Post-LASSO\, or \regularization parameter\.

Return your output in strict JSON format only, with no explanatory text outside the object.

RA Counts Prompt:

You are an expert assistant specializing in analyzing economics research papers for an audience of academic economists. Your task is to identify the number of research assistants (RAs) acknowledged in each paper. This information is located in the acknowledgment footnote (usually marked with an asterisk * on the first page). The paper will usually list a string of names followed by provided excellent research assistance or we thank [Name1], [Name2], and [Name3] for excellent research assistance, you must count every individual name in that list. Only count individuals explicitly tied to the phrases research assistance, RA, help with the data, assistance with the analysis, or coding support. Do not include anyone thanked for helpful comments, suggestions, discussions, or feedback.

- Here are some examples: “Arkey Barnett, Maria Canals, Madhavi Jha, Paola G. Villa Paro, Joshua Thomas, Lucía Valdivieso, and Qianyao Ye provided outstanding research assistance”, “We also thank John Dalton, Kevin Wiseman, and Jack Rossbach for extraordinary research assistance”, “We also thank Melanie O’Gorman and Lijun Zhang for excellent research assistance”.

Return results in a csv file, with the paper title, journal, category and subcategories, the year the paper was published, and an additional column titled “num_research_assistants”. Next, reread the acknowledgments for any paper with more than 5 research assistants. Carefully parse sentences with multiple clauses. For example, in the sentence: We thank [Name A] and [Name B] for comments, and [Name C] for research assistance”, only one name is an RA. If the acknowledgment is ambiguous (we thank X, Y, and Z for their help), do not count them. Return a table with the reverified count for these papers only.