Data Types and Empirical Methods in 27 Top Economics Journals*

Qu Feng, Zhifeng Wang, Shihao Zhou

March 13, 2026

Abstract

This paper examines the relevance and frequency of econometric methods in 3,125 empirical papers published in 27 top economics journals during 2014-2018. There are three major findings. First, 93% of these empirical papers use microdata, while only 7% and 4% of them employ macrodata and time series data, respectively. Second, two-stage least squares and difference-in-differences are the two leading identification strategies, accounting for nearly half and a quarter, respectively, in empirical studies using microdata. Third, there is a clear divide in empirical studies using microdata and macrodata in terms of data collection, identification strategies and econometric methods.

Keywords: Data Types, Empirical Methods, Identification Strategies

JEL Classification: C18; B41; C40

*Feng is from Economics Division, School of Social Sciences, Nanyang Technological University. 48 Nanyang Avenue, Singapore, 639818. Email: qfeng@ntu.edu.sg. ORCID ID: 0000-0003-1306-9632. Wang is from Li Anmin Institute of Economic Research, Liaoning University, China. Email: wang0869@e.ntu.edu.sg. ORCID ID: 0000-0003-0654-9953. Zhou is from School of Business, Singapore University of Social Sciences. Email: shzhou@suss.edu.sg. ORCID ID: 0000-0003-0015-7190. All authors are in alphabetical order and contributed equally to the paper and could be treated as the joint first author. Excellent research assistance from Tingbin Bian and Bowen Yang is greatly appreciated. Earlier versions of this article were presented at Singapore Economic Review Conference 2022; 2024 International Conference on Econometrics with Its Application and Practice in Finance at Nankai University, China; 2025 International Symposium on Time Series and Panel Data in the AI Era at Xiamen University, China; The 14th International Symposium of Quantitative Economics 2025 at Jilin University, China. Constructive comments from Qiang Chen, Ying Fang, Cheng Hsiao, Yongmiao Hong, Jiti Gao, Whitney Newey and the audience at these conferences are highly acknowledged. Feng acknowledges the financial support of the MOE Tier 1 grant RG110/21 from Nanyang Technological University.

1 Introduction

Over the decades, econometrics has advanced tremendously in various streams, including time series/ financial econometrics, panel data econometrics, microeconometrics, Bayesian econometrics, and nonparametric econometrics, among others. On one hand, given the wide range of econometric methods available and the limited time PhD students in economics and business schools have, it is natural for them to wonder how best to allocate their efforts. Given the wide range of available econometric methods, students typically develop a broad understanding of commonly used methods and may subsequently focus on techniques that aligned with their research interest, data types, and empirical challenges. At the same time, from the perspective of economics faculty, the structure and sequencing of econometrics courses often take into account both theoretical foundations and their relevance in empirical research. In addition, given the big pool of empirical methods, there is growing interest in understanding their prevalence in recent empirical studies. On the other hand, as pointed out by Heckman (2001), the gap between econometric theory and empirical economic research has been widening. He criticizes that econometrics has become too closely tied to mathematical statistics and has become less relevant to empirical research. Thus, to address these important questions and concerns, we look at the demand side of econometric methods in economics, and provide a data-driven approach to examine the popularity of them and identification strategies in the recent empirical economic studies published in top economics journals.

In this paper, we manually collect information on 3,125 empirical papers out of a total of 5,589 papers published in 27 top economics journals from 2014 to 2018.[1] Given the stunning fact that most articles in these top journals are empirical, we first look at the frequency of data types and data collection categories in these empirical papers. Specifically, we divide them into two types: microdata and macrodata. Following Cameron and Trivedi (2005, p.18), we define microdata as cross section or micro (or short) panel datasets used in empirical papers, whereas macrodata includes time series or macro (or long) panel datasets.[2] Unlike existing literature, this paper uses data type as the primary determinant for categorizing econometric methods used in empirical papers. Additionally, we also investigate the identification strategies in empirical papers using microdata,

[1] We select both top-tier general-interest journals and leading field journals that publish both theoretical and empirical research in our top 27 journal list. These journals are all ranked as A or above in the Australian Business Deans Council (ABDC) ranking, or are listed among the top 100 journals in the IDEAS/RePEc rankings. We acknowledge that our list is not exhaustive and that journals not included should not be regarded as less prestigious or significant in the field. Future research may also extend to finance journals and inference methods. We follow the definition of empirical papers in Angrist et al. (2017) and exclude methodological papers with empirical illustrations.

[2] Our definition of microdata is also consistent with the definition of microdata by the World Bank: https://datahelpdesk.worldbank.org/knowledgebase/articles/228873-what-do-we-mean-by-microdata

such as two-stage least squares (2SLS) (or instrumental variable (IV) estimation), difference-in-differences (DiD), generalized method of moments (GMM), regression discontinuity design (RDD), randomized controlled trials (RCT), and matching methods. Finally, by grouping the empirical papers by data types, we explore the frequency of the econometric methods in different journal groups and provide a ranking for them. The frequencies of nonparametric, quantile regression and Bayesian estimation methods across different journal groups are also recorded.

Our results contain three main findings. First, 93% of the empirical papers in the 27 top economics journals rely solely on microdata, while only 7% utilize macrodata and 4% use time series. Moreover, publishing in top journals in recent years often requires access to restricted or self-collected data rather than publicly available data. Second, 2SLS and DiD are the two most widely used identification strategies, accounting for nearly half and a quarter, respectively, of empirical studies using microdata. For example, 55 (and 94) papers employ DiD in the *American Economic Review* (AER) (and top 5) during 2014-2018, accounting for 30% of all empirical papers using microdata with a clear identification strategy. Third, there is a clear divide in empirical studies using microdata and macrodata in terms of data collection, identification strategies and econometric methods. For example, nonparametric methods and quantile regression are predominantly applied to microdata, whereas Bayesian approaches are more commonly used in studies employing macrodata.

In addition, we also document specific econometric methods by frequency in the 27 top economics journals. In particular, we observe that in the papers using microdata, the five most popular econometric methods besides OLS are panel data models, 2SLS, Logit/Probit models, maximum likelihood estimation (MLE), and GMM. However, for papers using macrodata, the most frequent five methods besides OLS in the top 27 journals are panel data models, vector autoregression (VAR) type, Bayesian, 2SLS, and autoregression (AR) type methods. Given that 63% of empirical papers in these 27 journals use panel datasets, it is not surprising that panel data models appear most frequently in empirical studies.

These findings may have important implications for us to learn, teach and research econometrics. For example, exposure to a broad range of econometric methods is common in graduate training, although the specific techniques applied in practice depend on the research topic and the nature of the data. When multiple methods are feasible for a given dataset and research question, patterns of methodological prevalence in the literature may provide insight into current methodological trends. Furthermore, we also provide rankings of specific econometric methods by frequency used

in different journal groups.

Nevertheless, there are several important considerations when interpreting our results. First, the frequency of econometric methods identified in our study does not necessarily reflect their overall importance or superiority. For example, nonparametric and semi-parametric methods remain among the top 10 techniques used in microdata papers within the top 27 journals, even if they appear less frequently than some other methods. In other words, we emphasize that all econometric methods are valuable, and researchers should not choose which methods to learn or apply solely based on publication prospects. Second, our analysis covers the period from 2014 to 2018, before the growing frequency of machine learning methods and recent econometric advances, such as heterogeneous-effect DiD models. It is important to note that our objective is to provide a retrospective perspective rather than to predict future research trends.

The remainder of the paper is organized as follows. In Section 2, we highlight our contributions to the existing literature. Section 3 describes the construction of our sample and methodology, whereas Section 4 presents our main findings. Lastly, Section 5 concludes. Additional results are included in Appendix 2.

2 Literature Review and Contributions

There are several important papers in the literature. Angrist et al. (2017) apply machine learning algorithms to 134,892 papers published in 80 economics journals from 1980 to 2015, classifying them into 10 economic fields. Their focus is on weights of the top five journals and *Review of Economics and Statistics* (ReStat), field shares of papers over time and the increasing fraction of empirical papers both within each field and across fields. By investigating all articles in the top five economics journals during 1970-2012, Card and DellaVigna (2013) find an increasing trend in the number of submissions, the length of articles, and the number of authors. In addition, the share of papers published in AER also increases from 25% in 1970 to 40% in 2012. On the other hand, Ellison (2002) examines the publication process of papers published in the top five economics journals and ReStat from 1970 to 1999, and finds a substantial slowdown of the publication process due to more demanding revisions.

Besides, the frequency of panel data regressions has also been extensively discussed in the literature. De Chaisemartin and d'Haultfoeuille (2024) report that 26 of the 100 most-cited papers in the AER from 2015 to 2019 estimate a two-way fixed effects (TWFE) model. However, De Chaisemartin and d'Haultfoeuille (2023) also highlight growing concerns about DiD analysis with

heterogeneous treatment effects, which TWFE cannot consistently estimate. Similarly, Arkhangelsky and Imbens (2024) review advances in causal panel data methods, warning that TWFE relies on overly strong assumptions, and emphasize alternatives like DiD and synthetic control approaches.

Recently, Galiani, Gálvez and Nachman (2025) employ machine learning techniques to examine how theoretical and applied economics fields interact with each other and with other fields from the perspective of citations based on 24,273 articles published between 1970 and 2016. Moreover, Garg and Fetzer (2025) distinguish between general claims and those documented via causal inference methods in 44,000 NBER and CEPR working papers from 1980–2023 using a custom language model, and document a substantial rise in the share of causal claims.

Our work closely relates to the following four papers. Hamermesh (2013) examines the coauthorship patterns, age structure of authors and data collection in the papers published in AER, *Journal of Political Economy* (JPE), and *Quarterly Journal of Economics* (QJE) during the 1960s-2010s. He observes a declining trend of using publicly available data. Panhans and Singleton (2015) focus on quasi-experimental methods and IV approach using the information of titles and abstracts of empirical papers in the top four economics journals (excluding the *Review of Economic Studies* (RES)) from 1970 to 2014 and 11 top field journals from 1990 to 2014. Xiao et al. (2014) summarize 9 main econometric methods in reduced form papers and popular identification strategies, including natural experiment, field experiment, DiD, matching and RDD in empirical papers using microdata in the top five journals during 2001-2012. Lastly, Goldsmith-Pinkham (2024) analyzes the evolution of the credibility revolution using a sample of 28,397 NBER working papers from 1982 to 2024, covering all fields of economics. In particular, he documents the rise of DiD methods in fields such as finance and macroeconomics over the past two decades.

Our paper adds to the existing literature in several ways. First, our paper is the first to group empirical papers by data type using microdata and macrodata and a more traditional definition of cross section, time series, and panel data. We use the prominence of microdata in empirical papers as a basis to categorize econometric methods. Second, instead of concentrating on the top five journals, our study covers empirical papers in 27 top economics journals in more recent years, such that our results can give a complete picture of empirical papers in all top economics journals. In particular, we distinguish between 14 journals focused on microeconomic topics, four on macroeconomic topics, and seven journals where the focus is primarily on non-structural approaches or reduced-form analysis. Third, our study examines the empirical methods used in the top 27 journals, with particular focus on panel data techniques. Additionally, we categorize the types of

data collection and document identification strategies. Overall, we record the frequency of over 40 econometric methods across different data types in top economics journals, including approaches such as nonparametric regression, quantile regression, and Bayesian methods.[3]

3 Sample Journals, Issues and Methodology

We manually collect data on 3,125 empirical papers published in 27 top economics journals from 2014 to 2018. These include the top five journals, five general interest journals, 10 top field journals, and seven additional top field journals.[4] Since the focus is on empirical research, top economics journals mainly publishing theoretical papers, such as, *Journal of Economic Theory*, *Theoretical Economics*, *Journal of Econometrics* and *Review of Economic Dynamics*, are excluded.[5] We collect all issues of 12 journals, including top five economics journals, five general interest journals, and *Journal of Labor Economics* (JLE) and the *RAND Journal of Economics* (RAND). Following the sampling approach in Hamermesh (2013), we randomly select half of the issues for the rest 15 top field journals.[6]

Inspired by Angrist et al. (2017), to explore different patterns in data types, identification strategies and econometric methods used in empirical papers in different research fields and journals, we classify all empirical papers into several journal groups, including Top 5, Top 27, General 4, Macro 4, Micro 14. Besides the Top 5, the remaining 22 journals are divided into three groups: General 4, Macro 4, and Micro 14. The General 4 group includes EJ, IER, JEEA and ReStat. Given the fact that empirical papers published in QE mainly use microdata, it is grouped to the Micro 14, which refers to journals publishing papers more related to microeconomic topics with microdata. Besides QE, they include JLE, RAND, JDE, JPubE, AEJ: Applied, AEJ: Policy, AEJ: Micro, JEEM, JHE, JHR, JIndE, JUE and JPopE. Correspondingly, the Macro 4 refers to four

[3]For definitions of the empirical methods mentioned in this paper, please see Table A6.

[4]For more details, please refer to the Table A1 in Appendix 2. Column (1) indicates journal names, including AER, ECTA, JPE, QJE, RES, EJ, IER, JEEA, QE, ReStat, JLE, RAND, JDE, JIE, JME, JPubE, AEJ Applied, AEJ Policy, AEJ Micro, AEJ Macro, JEEM, JHE, JHR, JIndE, JMCB, JUE, JPopE. Columns (2) and (3) list the total and collected issues of each journal. Column (5) shows the number of papers in each journal. In column (6), we also list the journal groups each economic journal belongs to.

[5]Except *Journal of Industrial Economics* and *Journal of Population Economics* listed as A, the rest 25 top economics journals we choose are categorized as A* journals by the Australian Business Deans Council (ABDC) list, with very few top empirical economics A* journals in the list are left out from our choice. https://abdc.edu.au/abdc-journal-quality-list/ In addition, 20 out of our 27 top journals are among top 40 economics journals in the RePEc journal list, which also includes journals in finance, business and those mainly publishing theoretical papers. https://ideas.repec.org/top/top.journals.all.html

[6]We reviewed a total of 5,589 papers from 27 leading economics journals, including non-empirical studies. Among these papers, there are 2,182 papers in the 15 journals with half coverage. Given the random sampling and the large number of papers reviewed, we believe that extending the sample coverage for these 15 top field journals to 100% coverage would not significantly alter the main results.

journals traditionally focusing on macroeconomic research topics, including JME, AEJ: Macro, JMCB and JIE. Similarly, seven non-structural journals are defined as a subgroup of Micro 14 which mainly rely on reduced form approaches and rarely use structural estimation.[7]

In each of these 3,125 empirical papers, we collect detailed information on data types, data collection categories, identification strategies, specific estimation methods and whether structural estimation or reduced form is applied. First, by data types, these empirical papers are grouped into papers using microdata *only* and those using macrodata.[8] Some papers use both micro-level and macrodata sets. To be consistent with traditional definition of data types, we also categorize these empirical papers into time series, cross section, and panel data groups, respectively.

Second, data collection categories are examined in these papers: publicly available sources, and supervisory data. The supervisory data includes restricted data collected or purchased from government authorities or data vendors, or data collected by field survey, field experiment, lab experiment, and web scraping algorithm, respectively. Third, we look at identification strategies used, including 2SLS (or IV), GMM, DiD, RDD, RCT, and matching.

Fourth, the applications of nonparametric, Bayesian and quantile regression methods in these empirical papers are also counted. Whether structural estimation or reduced-form approach is adopted in an empirical paper is also recorded. Finally, we record all the major empirical methods used in our sample, e.g., MLE, Probit, Heckit, panel methods for papers using microdata, panel methods, VAR, factor models for papers using macrodata.

4 Main Findings

In this section, our main findings on data types, data collection categories, identification strategies, and specific estimation methods are presented. In particular, we highlight similarities and differences among papers across different journal groups.

[7]Structural estimation is counted when it is explicitly mentioned and verified manually in papers reviewed. See Table A2 in Appendix 2 for the comparison across Top 5 journals. Based on the comparison of the top five journals, we observe similar patterns as those drawn from the full sample of all 27 journals. In particular, journals like *Econometrica* (ECTA) and the *Review of Economic Studies* (RES) show lower empirical ratios, at 23% and 33%, respectively, compared to the other three journals. Furthermore, while microdata appear in the majority of empirical papers in the top five journals, papers using macrodata account for a relatively small share. Also, structural estimation is most prevalent in ECTA (38%) and the RES (27%), with the QJE having the lowest proportion (6%). Additionally, publicly unavailable data is commonly used in the QJE, and reduced-form identification strategies are widely employed in ECTA.

[8]As noted in the previous paragraph, following the literature and the World Bank, we define microdata as cross sectional or micro (short) panel datasets, while macrodata refers to time series or long panel datasets.

4.1 Data Types

Table 1 summarizes the patterns of data types used in empirical papers in Top 27 economics journals from 2014 to 2018. Our analysis is based on the division of microdata only and macrodata, as well as traditional classification of cross section, time series and panel data. Some papers could use multiple datasets and data types, such as, both microdata and macrodata or both cross section and panel data. We present the results by seven journal groups in Table 1, with ratios in the upper panel and number of papers in the lower panel.

⟨Insert Table 1 here⟩

There are three major findings from the table. First, the full sample results in column (1) show that 93% of empirical papers use microdata (cross section and micro or short panels) only, reflecting a strong emphasis on individual-level data.[9] In contrast, only 7% of them use macrodata, which includes time series and long panel data. In particular, only 4% empirical studies use time series data. Notably, from columns (1) to (4), we find little difference across Top 27, the AER, Top 5 and four general interest journals. Second, when examining the distribution of cross sectional, time series, and panel data, micro panel data account for the largest share in empirical research, representing 60% of the 3,125 empirical papers (1,860). Additionally, 34% of papers, i.e., 1,056, use cross section data.

Third, as shown in the fifth row of Table 1, economic models play a critical role in empirical papers using macrodata in AER and Top 5 journals. 70% and 56% of empirical papers using macrodata in the AER and Top 5 journals employ economic models, compared to only 44% and 47% in papers using microdata only, respectively. This suggests greater reliance on models for interpreting macrodata in top journals.

4.2 Data Collection Categories

In Table 2, we categorize all empirical papers into three groups by their data collection methods: publicly available, restricted and self-collected data. The self-collected category includes four subcategories: field survey, field experiment, lab experiment, and Internet related.[10] Restricted data is from either confidential sources or purchased from external data vendors. Survey designed and conducted by external institutions also falls under this category.

[9] This finding is consistent with Table 9 by Glandon, et al. (2023) who focus on macroeconomic research papers published in several leading field and top 5 journals during 2016-2018.

[10] See Table A8 for the definitions of data collection categories.

⟨Insert Table 2 here⟩

First, column (1) of Table 2 shows that 52% of empirical papers throughout all Top 27 journals rely on publicly available data. Particularly, this percentage drops to 45% in the Top 5 journals, as shown in column (3), underscoring the increasing importance of restricted and self-collected datasets in top economic research. Column (3) of Table 2 also indicates that 34% of empirical papers in the Top 5 journals employ restricted data.

Second, based on column (3) of Table 2, field experiments are another key data source, accounting for 10% of papers in the Top 5 journals, while the use of internet-related data remains only 1% in all Top 27 journals on average. Interestingly, self-collected data, including field survey, lab experiments and internet related data, are rarely used in the four macro-topic journals.

Third, as shown in the first row of Table 2, there are more empirical papers in Top 27 journals on average, AER and Micro 14 journals, where they account for 56%, 60% and 66%, respectively. On the contrary, more theoretical papers appear in the other 4 of Top 5 journals (i.e., ECTA, QJE, JPE and RES) and Macro 4 journals. This is in line with the findings in Angrist et al. (2017).

To further explore differences in data types and collection categories between the Top 5 and other journals, we report the following regressions in Table 3:

$$y_i = \beta_0 + \beta_1 Top5_i + \varepsilon_i,$$

where $Top5_i$ is a dummy for Top 5 journal group.

First, Table 3, column (1) indicates that papers in the Top 5 journals are more likely to use restricted data, with publicly available data appearing 8.8% less frequently compared to other journals. Columns (2) and (3) document a decline in the use of publicly available data and an increase in restricted data over time. This pattern is consistent with Hamermesh's (2013) finding based on randomly selected papers in AER, QJE and JPE during the 1960s-2010s. Furthermore, columns (4) and (5) suggest that the Top 5 journals are associated with slightly greater reliance on time series data (by 1.5%) and on structural estimation methods (by 10.8%) compared to the other 22 journals in the sample.

On the other hand, the increasing popularity of restricted data in Top 5 journals may be related to the nature of the data itself. As indicated in Table 1, microdata comprises 92% of empirical papers in these journals, while macrodata accounts for only 8%. Therefore, the findings in Table 3 suggest an association between the use of microdata and the use of restricted datasets, whereas macrodata is more often associated with publicly available sources.

We observe a similar pattern in the Macro 4 and Micro 14 journal groups. As shown in Table 1, papers in the Macro 4 group rely on microdata in 65% of cases, while macrodata accounts for 35%'s the highest proportion among all journal groups. In contrast, nearly all papers in the Micro 14 group employ microdata (98%), with only 2% using macrodata. Consequently, Table 2 reports that the Macro 4 journals have the highest share of publicly available data (69%) across all groups. However, in the Micro 14 group, only about half of the papers (51%) rely on publicly available data. This suggests a correlation between the category of data used and the research focus of each journal group.

⟨Insert Table 3 here⟩

4.3 Identification Strategies

Identification strategies are essential to empirical studies and also the counts of some specific strategies in a subset of journals considered here have been recorded by Singleton (2015) and Xiao et al. (2014). Following Lewbel (2019), we further elaborate the relative importance of different identification strategies in papers using microdata in Table 4.

⟨Insert Table 4 here⟩

First, we observe that half of the papers employing microdata in the Top 27 journals have clear identification strategies in Table 4 column (1). In particular, 56% of the Top 5 journal papers demonstrate clear identification strategies, compared to only 24% in the four macro-topic journals, as shown in columns (3) and (5) of Table 4.

Second, we find that 2SLS (or IV) and DiD are the most popular identification strategies, accounting for 47% and 26% of these 1,550 papers.[11] The use of DiD is even higher in the Micro 14 and non-structural journals with ratios of 30% and 35%, respectively. Notably, combining 2SLS and GMM is particularly prevalent in the Top 5 journals, utilized in nearly 60% of papers.

Third, additional methods such as RDD, RCT, and matching are used in 10%, 6%, and 3% of empirical papers utilizing microdata in the sample. An interesting fact is that RDD and RCT are rarely used in 4 macro-topic journals.

In recent years, researchers such as Sun et al. (2025) have extensively explored DiD and synthetic control methods. In Table 5, we first present the increasing prevalence of DiD over time within the Top 27 journal group and a moderate increase in the Top 5 journals during 2014-2018, consistent

[11] See Table A9 in Appendix 2 for the full definitions of the identification strategies and empirical methods.

with the findings by Goldsmith-Pinkham (2024) based on the sample of NBER working papers. However, we do not observe the same trend in the published papers for synthetic control methods during this period, as reported in that study. This could be explained by the fact that synthetic control methods are mainly designed for empirical studies using macrodata, which account for only 7-9% of total empirical papers in the sample. Similarly, we do not observe a clear increasing trend of other identification methods such as RDD, GMM, and RCT in both the Top 27 and Top 5 journal groups in 2018.

⟨Insert Table 5 here⟩

4.4 Three General Methods: Nonparametric, Quantile and Bayesian

Table 6 reports the applications of three general methods in different journal groups: nonparametric, quantile regression, and Bayesian estimation. In addition, we also report the frequency of structural estimation used in the sample. First, column (1) of Table 6 suggests that structural estimation only contributes to 12% of the empirical papers in Top 27 journals, while the majority (88%) rely on reduced-form approach. In particular, the ratio of structural estimation is much higher in the Top 5 (20%), the AER (22%) and the Macro 4 journals (19%).

⟨Insert Table 6 here⟩

Second, as seen in the data collection and identification strategies above, there is a clear divide in empirical studies using microdata and macrodata in terms of econometric methods. The nonparametric and quantile regression methods are mainly for microdata. For instance, in the Top 27 journals from 2014 to 2018, only three papers use nonparametric methods, and only two apply quantile regression with macrodata. This highlights the limited role of these two methods in papers using macrodata with a relatively small sample size. Similarly, Bayesian methods are mainly used in empirical papers utilizing macrodata with a ratio of 83% in column (1), demonstrating a greater relevance in macroeconomic research. In contrast, they are underutilized in the Micro 14 journals, with only three out of 1,387 papers employing Bayesian methods, as shown in Column (6).

Third, in general, nonparametric (including semiparametric), quantile regression, and Bayesian estimation methods are not well applied in empirical papers, covering only 2%, 1%, and 2% in the Top 27 journals, respectively. In addition, papers in the Top 5 journals utilize slightly more nonparametric methods (5%), compared to the full sample. Again, the caveat here is that our

retrospective perspective based on these published papers does not reflect the latest development and future research trends.

Overall, structural estimation, Bayesian methods, and nonparametric approaches are observed less frequently in empirical economic research, whereas reduced-form estimation appears more prevalent in the literature. This trend indicates a continued preference for simpler, more widely understood methodologies in most top journals.

4.5 Frequency of Econometric Methods

We have explored the data types and identification strategies so far. Given the vast advancement of econometric methods in the past decades, it would be important to examine their relevance from the perspective of their applications in empirical studies. Table 7 reports the ten most frequently used econometric methods in papers published in the Top 27 journals, the AER, and the Top 5 journals, ranked by frequency.[12] Again, we group each of them by data types. The detailed definition of the econometric methods can be found in Table A9 in Appendix 2. For example, 2SLS in this paper refers to a class of methods, including IV estimation, 2SLS, 3SLS, control function approach, simultaneous equation model and Bartik estimator.

⟨Insert Table 7 here⟩

First, comparing the Top 27, AER and the Top 5 papers using microdata, we observe that the most popular five econometric methods besides OLS are the same: panel data models (1,859 out of 2,900 papers in the Top 27), 2SLS (727), Logit/Probit models (455), MLE (186), and GMM (122).[13] The next leading methods used for microdata in the Top 27 journals are Poisson, Tobit/ Heckit, and nonparametric.

Second, within the Top 27 journals, 2,900 papers employ microdata, while 225 papers utilize macrodata. Among studies based on microdata, the most commonly used econometric methods are panel data models (1,859 papers), followed by 2SLS (727), Logit/Probit models (455), MLE (186), and GMM (122). For papers using macrodata, panel data models are the most frequently used

[12] Table A4 in Appendix 2 presents the full ranking, including all listed methods used in these journal groups, whereas Table A7 shows the results of another four journal groups.

[13] OLS in this paper refers to a class of methods, including ordinary least squares, general least squares (GLS), feasible general least squares (FGLS), weighted least squares (WLS), Seemingly Unrelated Regression (SUR), semi-log linear model and linear probability model, DOLS and FMOLS. Besides, fixed effects using panel data are also categorized under OLS. Similarly, MLE type methods include It includes MLE, simulated MLE, pseudo MLE, Beta/Gamma pseudo MLE, quasi MLE, fixed point MLE, full information MLE, truncated MLE, limited information MLE, restricted MLE, multiple indicators and multiple causes (MIMIC) model, and latent variable MLE.

(123 papers), followed by VAR models (84), Bayesian methods (48), 2SLS (39), and AR models (25).

Third, the ranking of the 10 leading econometric methods in the Top 5 journals differs slightly from that in the Top 27 journals. While the top five econometric methods used in the Top 5 journals employing microdata follow the same patterns as those in the broader Top 27 journals, the subsequent five methods appear in a different order. Specifically, nonparametric methods are more popular in the Top 5 journals, appearing in 33 out of 673 microdata papers, followed by Poisson models (23 papers), SMM (21 papers). For macrodata papers in the Top 5 journals, the most frequently used methods are panel data models (25 out of 60 papers), followed by VAR models (20), Bayesian methods (17), 2SLS (13), and SMM (5). The use of other econometric methods in macrodata studies within the Top 5 journals is negligible.

Similarly, the top five econometric methods in AER papers utilizing microdata follow the same pattern, with panel data models used in 63% of papers (200 out of 320), followed by 2SLS (25%), Logit/Probit models (12%), MLE (6%), and GMM (4.6%). The next most common methods are nonparametric techniques (4.3%), Poisson models (4%), SMM (3.4%), and minimum distance estimation (MDM) (3.4%). On the other hand, for AER papers employing macrodata, Bayesian methods are the most popular, appearing in 35.5% of papers (11 out of 31), followed by panel data models (34%), VAR models (29%), SMM (13%), and maximum likelihood estimation (9.6%). The subsequent leading methods are AR models (9.6%), 2SLS (9.6%), GMM (6.4%), and NLS (6.4%).

To further link econometric methods to data types, Tables 8 and 9 report the ten most frequently used time series and panel data methods in papers using macrodata, respectively.[14] Among all the Top 27 journals as well as Macro topic journals, VAR model is the most popular econometric method, followed by Bayesian approaches. On the other hand, Bayesian estimation is more popular among the AER and Top 5 journals, with VAR being the second most preferred methods in these two journal groups. Additionally, SMM is a preferred approach in papers published in AER, whereas 2SLS method stands out for its frequency in the Top 5 economics journals.

⟨Insert Table 8 here⟩

In addition, Table 9 presents a list of top ten methods employed in macro panel data papers across various journal groups. Based on the Top 27 column, we first find that Fixed Effect (FE) and 2SLS are the two most important methods in those journals, aligning with existing literature

[14]Tables A5 and A6 in Appendix 2 present the full ranking, including all listed methods used in these journal groups

(De Chaisemartin and d'Haultfoeuille, 2024). The rest of the columns exhibit a similar pattern. Conversely, MLE, along with the VAR model and its variations such as Structural VAR (SVAR) and Panel VAR (PVAR) models, are prevalent in both the Top 5 and AER journals. Next, we find that Bayesian is ranked among the top five in Macro topic journals, while their applications are limited in other journal groups.

⟨Insert Table 9 here⟩

We also report the patterns among the papers across different journal categories. Meanwhile, we present the time-varying patterns in Appendix 2.[15]

5 Conclusion

In this paper, we manually collect information on the data types and econometric methods in 3,125 empirical papers published in 27 top economics journals during 2014-2018. First, we document the widespread adoption of microdata and the increasing use of restricted and self-collected data in empirical studies. Second, we find that papers studying microdata frequently rely on identification strategies such as 2SLS and DiD, which are commonly applied for addressing endogeneity and constructing causal relationships. In contrast, for macrodata, theoretical models play a more significant role.

Third, the findings of econometric methods in the Top 5 journals further highlight the most widely adopted approaches. For microdata, the top five methods are 2SLS, GMM, Panel models, Logit/Probit models, and MLE. In contrast, for macrodata, the most frequently used methods include Panel models, VAR, Bayesian methods, 2SLS, and AR models. These findings may have implications for the coverage of econometric methods in graduate training, particularly given their prominence in leading economics journals.

Lastly, while techniques such as Bayesian methods, nonparametric models, and quantile regression are less commonly applied overall, their application varies across different research contexts. For instance, Bayesian methods are more prevalent in studies using macrodata, whereas nonparametric methods and quantile regression are more commonly observed in papers with microdata. These patterns may have implications for methodological coverage in econometrics training, particularly in areas where alternative approaches are less commonly employed.

[15] See Tables A3 in Appendix 2 for time-varying results. Overall, we observe similar patterns with some variations over time.

In conclusion, our paper documents the prominent use of microdata in empirical studies as well as restricted and self-collected data in top economics journals. These findings provide an overview of current methodological patterns in top-tier economic research. In addition, they may also have implications for the structure and content of econometrics training at the undergraduate and postgraduate levels, e.g., the course coverage suggested by Angrist and Pischke (2017, Table 2).

6 References

Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2017). Economic research evolves: Fields and styles. *American Economic Review*, **107(5)**, 293-97.

Angrist, J. and J. Pischke (2017). Undergraduate Econometrics Instruction: Through Our Classes, Darkly. *Journal of Economic Perspectives* **31(2)**: 125–44.

Arkhangelsky, D., and Imbens, G. (2024). Causal models for longitudinal and panel data: A survey. *The Econometrics Journal*, **27(3)**, C1-C61.

Card, D., and DellaVigna, S. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, **51(1)**, 144-61.

Cameron, A. and Trivedi, P, (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press.

De Chaisemartin, C., and d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, **26(3)**, C1-C30.

De Chaisemartin, C., and d'Haultfoeuille, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, **1-45**.

Ellison, G. (2002). The slowdown of the economics publishing process. *Journal of Political Economy*, **110(5)**, 947-993.

Galiani, S., Gálvez, R.H. and Nachman, I. (2025). Specialization trends in economics research: a large-scale study using natural language processing and citation analysis. *Economic Inquiry*, 63(1), 289–329.

Garg, P., and Fetzer, T. (2025), Causal claims in economics, working paper.

Glandon, P. J., K. Kuttner, S. Mazumder and C. Stroup. 2023. Macroeconomic Research, Present and Past. *Journal of Economic Literature*, 61 (3): 1088–1126.

Goldsmith-Pinkham, P. (2024). Tracking the credibility revolution across fields. *arXiv preprint arXiv:2405.20604*.

Hamermesh, D. S. (2013). Six decades of top economics publishing: Who and how?. *Journal of Economic Literature*, **51(1)**, 162-72.

Heckman, J. (2001). Econometrics and empirical economics. *Journal of Econometrics*, **100(1)**, 3-5.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, **95(1)**, 161-182.

Lewbel, A. (2019). The identification zoo: meanings of identification in econometrics. *Journal of Economic Literature*, **57(4)**, 835-903.

Panhans, M. T., and Singleton, J. D. (2017). The empirical economist's toolkit: from models to methods. *History of Political Economy*, 127-157.

Sun, Y., Xie, H., and Zhang, Y. (2025). Difference-in-Differences Meets Synthetic Control: Doubly Robust Identification and Estimation. *arXiv preprint arXiv:2503.11375*.

Xiao, J., Ren, F., and Liu, Y. (2014). Quantitative analysis of articles in major English economics journals. *World Economy* (in Chinese), **(1)**, 148-160.

Appendix 1: Coding Protocol

This appendix documents the procedures used to construct the analytical dataset from the raw file and to classify each paper based on key categories, including journal group, data type, data source, and empirical methods.

1. Sample Construction

The unit of observation is an individual paper. The sample is constructed by applying a sequence of filters based on journal group and publication year. Journal classifications are defined using indicator variables (e.g., top5, top22, gen4, macrojournals, microjournals), corresponding to the proposed journal groups. The baseline specification restricts the sample to a selected journal group (e.g., top5) and a specific year (e.g., year == 2018). To focus on empirical work, papers without empirical content are excluded based on an indicator identifying non-empirical studies (e.g., noofempirics). All reported statistics are therefore conditional on the set of empirical papers within the selected sample (e.g., empiricalpaper).

2. Coding of Data Types

Each paper may report up to two data-type fields (e.g., data_type1 and data_type2), indicating whether the data correspond to cross sectional, time series, or panel data. Papers are classified into these categories based on the information provided in these fields. A paper is classified under a given data type if any of the reported fields indicates the use of that data type. Accordingly, a paper may be assigned to multiple data categories where more than one type of data is utilised (e.g., crosssection_data, ts_data, and panel_data).

3. Data Source Categories

Data sources are classified to distinguish between different types of data (e.g., public data, field surveys, field experiments, laboratory experiments, internet-based data, and restricted data). Indicator variables (e.g., publicdummy, fieldsurveydummy, fieldexpdummy, labexpdummy, internetdummy, restricteddummy) are constructed to capture each category. For each category, total counts and corresponding rates are computed relative to the number of empirical papers in the sample (e.g., empiricalpaper).

4. Macro vs Micro Classification

A key distinction in the analysis is between â€œmacroâ€ and â€œmicroâ€ empirical work. This classification is primarily based on an indicator identifying whether a paper uses time series or long panel data (timeseriesorlongpanel1foryes0for). To ensure consistency, this indicator is cross-checked against the data type fields. In cases where the original indicator is missing or incomplete, but the paper is identified as using time series or long panel data based on the data type information, it is classified as macro paper. Papers not classified as time series or long panel are treated as micro paper. Panel data are further divided into â€œmacro panelâ€ and â€œmicro panelâ€ according to this classification.

5. Coding of Empirical Methods

Empirical methods are recorded across multiple fields (e.g., empirical_method1-6 for micro methods, and macro_method1 to macro_method7 for macro methods). These fields are mapped into a set of standardised method categories (e.g., OLS, 2SLS, DID, RD, GMM, VAR, Bayesian, and factor models). For each method category, a paper is classified as using that method if it is reported in any of the corresponding fields. Accordingly, a paper may be associated with multiple method categories where more than one empirical approach is employed. Method usage is aggregated by summing the corresponding indicators across papers. The rates are then calculated relative to the number of empirical papers (or macro papers, where appropriate).

6. Treatment of Ambiguous and Missing Cases

Several procedures are implemented to ensure consistent handling of ambiguous or incomplete information:

a. Multiple entries across fields

When multiple data types or methods are listed for a paper, all relevant categories are retained. For example, a paper is classified as time series if either data_type1 or data_type2 indicates time series data, and method indicators such as olsdummy are defined based on whether the method appears in any of microdummy1â€“6.

b. Missing classification variables

When key indicators (e.g., timeseriesorlongpanel1foryes0for) are missing or incomplete, they are inferred from other available information to avoid misclassification. For instance, this indicator is reconciled with ts_data to ensure that papers identified as time series are consistently classified as such.

c. Standardisation of method labels

Closely related or overlapping method labels are grouped into broader categories to ensure consistency. For example, selection models (e.g., HECKIT) are grouped with Tobit models, while various time series approaches (e.g., AR, VAR, GARCH, ECM, and cointegration) are consolidated into a common category using variables such as microdummy1.

d. Interpretation of shares

Because papers may be assigned to multiple categories, reported shares should be interpreted as the proportion of papers that use a given method or data type, rather than as mutually exclusive partitions. For example, measures such as ts_lp_rate and ts_micro_rate are computed relative to the relevant sample and may overlap across categories.

7. Inter-Coder Reliability and Internal Consistency Checks

To ensure coding reliability and internal consistency, we conducted systematic cross-checks across method classifications. In particular, we verified the coding by summing papers across categories whenever the classifications were mutually exclusive. For example, we aggregated the number of microdata and macrodata papers in Table 1, as well as the number of papers under different identification strategies in Table 2, and confirmed that these totals matched the overall number of empirical papers. This procedure ensured that the classifications were applied consistently and that the grouping of related methodologies did not introduce double counting or other inconsistencies in the reported totals.