Introduction to the *Handbook of Research Methods and Applications in Empirical Microeconomics*

Nigar Hashimzade and Michael A. Thornton

Few areas in economics have seen the growth in interest and influence over the last two decades experienced by empirical microeconomics. The development of modern research methods in microeconomics has been driven, on the one hand, by the accumulation of numerous data sets at the level of individuals, households and firms, and, on the other hand, by an increase in the capacity and speed of computers. While the more traditional empirical methods were largely limited to establishing correlation among economic variables, the new methods aim to uncover causality. At the same time, progress in computing has enabled researchers to estimate complex relationships among variables suggested by economic theory, without resorting to crude approximation, or to use entirely data-driven estimation. These advances have opened new possibilities for applied research in microeconomics and many other fields in economics and social sciences, most notably, in the estimation of sophisticated structural models and in the evaluation of policy interventions.

This volume introduces a range of modern empirical methods with applications in microeconomics and some illustrations of the use of software packages. The first section, “Econometric Methods in Microeconomics”, presents a selection of linear, non-linear and semiparametric methods, with examples from political economy, labour economics and the analysis of auctions.

The focus of the opening chapter, by Ryo Okui, is on the estimation of linear dynamic panel models, common in applied microeconomics (Chapter 1). The chapter gives an overview of the difficulties arising in application of the standard generalised method of moments estimator and presents potential solutions. The empirical example demonstrates the use of this more advanced methodology in estimation of a relationship between economic environment and democracy. The next chapter, by Anna Gloria Billé, presents spatial autoregressive non-linear models (Chapter 2). These models are a natural extension of non-linear models to the empirical context where cross-sectional dependence can be related to relative proximity of cross-sectional units. The empirical example illustrates how the geographical coordinates of regions in Europe can be exploited in the analysis of the employment patterns across European countries. This is followed by a chapter by Tong Li and Xiaoyong Zheng that investigates semiparametric methods and partial identification approaches in the empirical analysis of the auction data (Chapter 3). Particular attention is paid to the role of the informational assumptions in the models. The concluding chapter in this part, by Martin Huber, introduces the reader to several flexible methods of policy evaluation, including randomised and quasi-random treatment, instrumental variables, difference-in-differences, regression discontinuities and data-driven approaches based on machine learning (Chapter 4). These methods are presented in more detail with examples in subsequent parts of this volume.

Each chapter in the second section, “Households, Businesses and Societies”, presents key empirical methodologies used in the analysis of contemporary economic, social and policy

issues. The wide variety of applications illustrated in this section includes econometric analysis of fertility, measurement of discrimination in the labour market, design and evaluation of tax-transfer systems, pricing strategies, performance analysis of banking and social epidemiology. The main focus of the chapters in this section is on a selection of statistical, structural, semi- and nonparametric methods.

The section opens with a chapter by Alfonso Miranda and Pravin K. Trivedi presenting several methods of estimation of a structural microeconomic model from cross-section and panel data, including the standard ordinary least squares, count models, hazard models and a discrete-time multi-spell duration model (Chapter 5). The advantages and shortcomings of each method are discussed in the context of their application to estimation of the socioeconomic and demographic determinants of fertility. This is followed by a chapter by Emmanuel Duguet reviewing several methods of empirical investigation of discrimination in the labour market, with a focus on discrimination in hiring and on wage discrimination (Chapter 6). The range of methods includes statistical and structural evaluation, such as the comparison tests, stacked regressions, component models and extended Blinder–Oaxaca decomposition. The focus of the next chapter, by Ugo Colombino, is on the structural empirical methods of policy evaluation with application to the evaluation and design of tax and transfer systems (Chapter 7). This chapter presents a range of structural approaches to the empirical analysis, such as the random utility maximisation, and labour supply responses to taxes and transfer at the extensive and the intensive margins. It concludes with an illustration of how structural microeconomic modelling can be combined with microsimulation and numerical optimisation. The next chapter, by Khai X. Chiong, Yu-Wei Hsieh and Matthew Shum, considers methods to construct bounds on counterfactual choice probabilities useful for counterfactual exercises in aggregate discrete-choice demand models, motivated by an empirical application using scanner data (Chapter 8). By taking a semiparametric approach, the analysis accommodates a wide variety of substitution patterns among alternatives while remaining a tractable linear programming problem. The next chapter, by Natalya Zelenyuk and Valentin Zelenyuk, expounds the current state of the art methodologies for the analysis of the performance of banks (Chapter 9). Beginning with the popular productivity and efficiency methods, data envelopment analysis and stochastic frontier analysis, the chapter reviews other methods whose popularity comes from the finance literature before changing focus to the issues of causal inference, such as differences-in-differences and regression discontinuity design, among others, that are the subject of the next section. This section concludes with a discussion by Christopher F. Baum of how the research methods commonly used in empirical microeconomics, for example to study disparities in income, wealth and human capital within and between societies, are becoming increasingly relevant to research in social epidemiology, particularly in understanding the origins of health disparities and analysing the policies used to address them (Chapter 10). As part of the discussion, this chapter highlights various techniques, including fixed-effects models, instrumental variables methods, difference-in-difference models, binary and ordered response models, count data models and conditional mixed-process models.

A detailed exposition of the modern methods of causal inference for policy evaluation and the statistical models underlying these methods is presented in the third section, “Policy Evaluation and Causality”. The approaches covered in this section include experimental and quasi-experimental methods, regression discontinuity designs, matching and difference-in-differences. The section begins with an overview chapter, by Denis Fougère and Nicolas Jacquemet, describing the main impact evaluation methods, both experimental and quasi-

experimental, and their underlying statistical models (Chapter 11). Regression discontinuity designs, which enable researchers to identify parameters of interest using discontinuities in the implementation of the policy across the population, have gained great popularity in policy evaluation in the recent years. This approach is the subject of the next chapter, by Otávio Bartalotti, Marinho Bertanha and Sebastian Calonico (Chapter 12). The chapter discusses key matters to consider for estimation, inference and overall implementation, such as local polynomial estimation, bandwidth choice and robust inference alongside recent extensions addressing common challenges, such as running variable manipulation, multiple cut-offs, measurement error and sample selection. The section concludes with a chapter by Emmanuel Duguet covering one of the most widely explored areas for policy evaluation,—the labour market (Chapter 13). The author uses evaluation of the impact of health events, for example, illness or accident, on labour market outcomes, such as labour market participation, to motivate a detailed exposition of the most common methods of estimation of the treatment effect in cross-sectional and panel data.

The fourth section, “Networks and Big Data in Microeconomics”, presents computationally intensive empirical methods. The applications illustrated in this section include the analysis of data generated by social media, econometric analysis of network and the machine-learning methods in causal inference.

The collection, visualisation and empirical analysis of the new flows of data that can be harvested from social media and web searching is the topic of the first chapter in this section, by Tho Pham, Piotr Śpiewanowski and Oleksandr Talavera (Chapter 14). This chapter presents a comprehensive overview of the tools and methods used in the economic literature, including discussion of the Stata and Python codes. These methods are illustrated in practical examples of the analysis of exchange rate responses to political uncertainty and labour market responses to the COVID-19 pandemic. Taking this topic further, the next chapter, by Pedro C.L. Souza, surveys the growing literature on the econometrics of networks when the network itself is either not observed or only partially so, this is to say when the meta-data concerning connections between nodes, or who is following or talking to whom, is not available (Chapter 15). It highlights recent research on the use of aggregate relational data and how connections might be inferred from observable (endogenous and exogenous) covariates. Next, Vishalie Shah, Noemi Kreif and Andrew M. Jones discuss the estimation of heterogeneous treatment effects across population subgroups based on their observed characteristics using three machine learning algorithms, the X-learner, the R-learner and causal forests (Chapter 16). These techniques are then illustrated in a policy impact evaluation of a national health insurance programme in Indonesia.

The concluding section showcases the use in empirical microeconomics of the two most popular software packages, Stata and R, with examples both of how to use existing routines and of how to write bespoke code. Applications include stochastic frontier analysis and estimation of the traditional and nested constant-elasticity-of-substitution (CES) function.

The first chapter in this section, by Oleg Badunenko, looks at stochastic frontier analysis using Stata (Chapter 17). As well as covering existing routines and packages, it includes a guide on coding new, bespoke commands using Stata’s own matrix programming language, Mata, with a particular focus on the analysis of panel data. The remaining chapters in the section discuss the use of the open-source statistical package, R. The topic of the first of these, by Giovanni Baiocchi, is the “ecosystem” of existing data processing R packages known as the *tidyverse*, that can facilitate econometric research workflows (Chapter 18). This wide-ranging chapter discusses the deployment of these existing resources for a wide array of popular

models, including generalised linear models, binary, count, multinomial, ordered, censored and truncated responses, fraction and proportion responses and applications in several fields of empirical microeconomics, including instrumental variables, panel data and quantile regression. The next chapter, by Giovanni Millo, concentrates on one of the most widely explored of those fields, introducing the mechanics of panel data econometrics with R before expanding upon the various robust estimators, which relax the restrictive hypothesis of error homoskedasticity, available in the *plm* package, and how to employ them in a microeconomic context (Chapter 19). The final chapter, by Arne Henningsen, Géraldine Henningsen and Gergő Literáti, explores the estimation of the CES function, which, despite being widely used across economics, is rarely estimated as it is not amenable to linear regression (Chapter 20). The authors show how to estimate CES functions, first with the traditional two arguments and then with three or more arguments, in R using the package *micEconCES*.

The intended readership is graduate students in economics and other social sciences as well as researchers looking for the most advanced methods of micro-level data analysis and policy evaluation. It is assumed that the readers are familiar with the core methods in econometrics and statistics for data analysis in economics, business and social sciences, at least at an intermediate level. For background reading on the core and advanced methods, along with the basics of Stata and R software, we recommend the excellent texts listed below, among others.

Lastly, we would like to express our gratitude to Professor Mark Casson for encouraging us to edit this volume, as well as our previous edited volume in this series, on research methods and application in empirical macroeconomics.

RECOMMENDED BACKGROUND READING

- Baum, Christopher F. (2016). *An Introduction to Stata Programming*. 2nd edition, Stata Press.
 Baltagi, Badi H. (2021). *Econometric Analysis of Panel Data*. 6th edition, Springer.
 Cameron, A. Colin and Pravin Trivedi (2021). *Microeconometrics Using Stata*. 3rd edition, Stata Press.
 Croissant, Yves and Giovanni Millo (2018). *Panel Data Econometrics with R*. Wiley.
 Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd edition, MIT Press.