# TOPICS IN IDENTIFICATION, LIMITED DEPENDENT VARIABLES, PARTIAL OBSERVABILITY, EXPERIMENTATION, AND FLEXIBLE MODELING: PART B

# ADVANCES IN ECONOMETRICS

Series Editors: Thomas B. Fomby,
R. Carter Hill, Ivan Jeliazkov,
Juan Carlos Escanciano,
Eric Hillebrand, Daniel L. Millimet,
Rodney Strachan,
David T. Jacho-Chávez and
Alicia Rambaldi

Previous Volumes:

Volume 29: Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill

Volume 30: 30th Anniversary Edition – Edited by Dek Terrell and Daniel Millmet.

Volume 31: Structural Econometric Models – Edited by Eugene Choo and Matthew Shum.

Volume 32: VAR Models in Macroeconomics — New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy

Volume 33: Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park

Volume 34: Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier.

Volume 35: Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman

Volume 36: Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee.

Volume 37: Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage, and R. Kelley Pace

Volume 38: Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano

Volume 39: The Econometrics of Complex Survey Data: Theory and Applications – Edited by Kim P. Huynh, David T. Jacho-Chávez and Guatam Tripathi

Volume 40A: Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A – Edited by Ivan Jeliazkov and Justin L. Tobias

ADVANCES IN ECONOMETRICS VOLUME 40, PART B

# TOPICS IN IDENTIFICATION, LIMITED DEPENDENT VARIABLES, PARTIAL OBSERVABILITY, EXPERIMENTATION, AND FLEXIBLE MODELING: PART B

EDITED BY

**IVAN JELIAZKOV**
*University of California, USA*

**JUSTIN L. TOBIAS**
*Purdue University, USA*


United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019



**British Library Cataloguing in Publication Data**
A catalogue record for this book is available from the British Library

ISBN: 978-1-83867-420-5 (Print)
ISBN: 978-1-83867-419-9 (Online)
ISBN: 978-1-83867-421-2 (Epub)

ISSN: 0731-9053 (Series)


ISOQAR certified Management System, awarded to Emerald for adherence to Environmental standard ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

# CONTENTS

# FOREWORD TO PART B

Part B of *Advances in Econometrics, Volume 40*, contains 11 chapters on topics that are distinct from those considered in Part A. Part B examines innovations in stochastic frontier (SF) analysis, nonparametric and semiparametric modeling and estimation, A/B experiments, and quantile regression. Together, the two parts form one of the largest volumes to date in the *Advances in Econometrics* series.

Two chapters in the volume address issues in SF models. The chapter by *Gholamreza Hajargasht and William E. Griffiths* considers a semiparametric SF model with correlated effects, where covariates in the production frontier enter nonparametrically. Specifically, they tackle a model of the form:

$$y_{it} = f(x_{it}) - u_i + v_{it},$$

where $u_i$ is a time-invariant one-sided inefficiency term and $f$ is modeled via penalized splines. They consider cases where $x$ is both univariate and multivariate. For the multivariate case, they take up partially linear, additive, and general nonparametric forms, treating the final case via multivariate spline bases. The methodology is illustrated in extensive Monte Carlo experiments. *Eri Nakamura, Takuya Urakami, and Kazuhiko Kakamu* consider a fully parametric, cross-sectional SF model with a greater focus on a specific application. They apply the SF model to examine how the division of labor (as measured by the number of firm sections) impacts total costs among a sample of 79 Japanese water suppliers in 2010. They leverage the population served by the $i$th supplier as a source of exogenous variation and describe inference via Markov chain Monte Carlo (MCMC). They find that ignoring the endogeneity of the number sections (or functional units) within a supplying firm provides a conservative underestimate of its effect on total cost, and that costs rise with the number of sections. Their work points toward a policy implication where the integration of internal divisions within Japanese water suppliers leads to more cost-effective production.

A set of three chapters examines topics in semiparametric and nonparametric modeling. A chapter by *Justin L. Tobias and Joshua C. C. Chan* presents an interesting modeling approach that aims to resolve an important difficulty in the modeling of unknown regression functions. Because such modeling typically nests the important linear case at the boundary of the parameter space, it can only approximate, but not reproduce exactly, the linear regression model. The chapter proposes a hierarchical setup in which a Bernoulli variable determines whether the model is exactly linear or nonlinear (for some predetermined thresh-

old level of nonlinearity), and hence in the course of sampling, the realizations of that variable can be used to infer the probability of each of the two cases. The methodology and its usefulness are demonstrated in generated data experiments and an application involving the impact of the body mass index on labor market earnings. *Jianghao Chu, Tae-Hwy Lee, and Aman Ullah* consider the analysis of semiparametric single index models in big data settings with many regressors, of which few may be relevant. For this reason, variable selection is a key concern in this setting. The authors propose the SIM-Rodeo algorithm, which handles variable selection as local bandwith selection, and shows its consistency in variable selection. A Monte Carlo study shows its superior performance relative to alternatives such as the SIM-Lasso. In an important chapter, *Edward George, Purushottam Laud, Brent Logan, Robert McCulloch, and Rodney Sparapani* pursue a fully nonparametric extension of the Bayesian Additive Regression Tree (BART) model. While BART achieves flexible modeling of the conditional mean function through ensambles of trees, the authors generalize its applicablity to the case of non-Gaussian errors by using Dirichlet process mixtures (DPM). An important motivation for the chapter is that estimation should be not only pursued without relying on strong assumptions but also in a fairly automatic way that does not require too much tuning in a wide variety of settings. The authors show that the resulting DPMBART model performs well in simulated and real data studies with normal and non-normal errors.

This volume also contains two chapters developing methodologies useful for making quick and reliable inference in so-called A/B experiments. These experiments often involve millions of observations and are conducted routinely by large online businesses. The goal of these experiments is to assess the effect of various treatments such as changes in webpage design on revenue; the difference between treated and untreated outcomes is commonly referred to as *lift*. *John Geweke* introduces BABI, or Bayesian analysis of A/B data. He accounts for the prevalence of zeros and heavy tails in such data by considering a zero-inflated log-normal model. A key advantage of BABI is its ability to allow for informed priors surrounding values of lift; Geweke specifically considers nine states related to lift *breadth* and *depth* corresponding to either no effect, the anticipated effect, or some other effect of treatment in each dimension. He applies this methodology using data from 21 different experiments conducted by a business with a large online presence, and finds that BABI yields tighter credible intervals than alternate approaches. *Hedibert Freitas Lopes, Matthew Taddy, and Matthew Gardner* consider the issue of inference of means of heavy-tailed distributions, a feature that often characterizes distributions of online expenditure data. Their approach is to specify the likelihood as multinomial with a finite (but large) number of support points for values less than some threshold and to allow for heavy tails by specifying the density above the threshold as a generalized Pareto distribution. They show how Bayesian inference can be conducted on the sampling weights and tail parameters using a novel independence Metropolis–Hastings algorithm. The results are applied to a large A/B experiment containing more than $10^7$ observations in eBay.

Three chapters in the volume relate to quantile regression. The chapter by *Mohammad Arshad Rahman and Angela Vossmeyer* discusses Bayesian techniques for estimating quantile regressions in binary longitudinal data settings. In Bayesian approaches, quantile regression typically employs the asymmetric Laplace distribution, noting the connection between the posterior mode under such a likelihood and a flat, improper prior, and the traditional frequentist quantile regression estimator. A useful additive mixture representation of the asymmetric Laplace, first noted by Kozumi and Kobayashi (2011) and skillfully adapted by Rahman and Vossmeyer to a binary longitudinal specification, yields computationally convenient MCMC inference. Rahman and Vossmeyer apply their methodology in two interesting applications: one involving determinants of female labor force participation and a second involving home ownership. *Debajit Dutta, Subhra Sankar Dhar and Amit Mitra* also consider the issue of quantile estimation, but of a location parameter in a stochastic volatility model. They derive the asymptotic distribution of the quantile estimator without assuming the density of the error is positive near the population quantile. They also discuss a Bayesian estimator based on the asymmetric Laplace likelihood. The third chapter in the set – by *Mohammad Arshad Rahman and Shubham Karnawat* – introduces a framework for flexible Bayesian quantile regression for ordinal outcomes. To overcome the problem that the skewness of the asymmetric Laplace distribution is determined by the chosen quantile, the authors study the generalized asymmetric Laplace (GAL) distribution. Their derivation of the GAL cumulative distribution and moment generating functions allows them to construct a likelihood function and posterior density that is explored by MCMC methods. The authors further demonstrate the advantages of their approach in an extensive simulation study and an application to public opionion on home ownership in the United States after the Great Recession.

The volume closes with "A Reaction," a brief comment from Dale J. Poirier on the June 2018 Advances in Econometrics Conference. Readers may take away from this short note a belief that the market for his classic *Intermediate Statistics and Econometrics* text has been surprisingly limited, and consumers of *Men Are from Mars, Women Are from Venus* should find his work equally appealing. We have all heard stories of successful marriages between devout Democrats and staunch Republicans, cat lovers and dog enthusiasts, and Michigan fans and Ohio State devotees. In two former students and contributors to this volume, Dale notes a similar warm stasis between a Bayesian and a frequentist. Whether or not the balanced presentation of his text served as a model for this otherwise inexplicably successful union of Hatfields and McCoys, the editors cannot say. We are tempted to recommend that Dale stick with his day job rather than branch out into couples therapy, but we also understand that his days are now free and unclaimed by the academy. Regardless of its true scope of influence, we remain grateful for his book and its many lessons, this opportunity to pay tribute to Dale, and the very positive impacts he has had on our careers.

# ACKNOWLEDGMENTS

We would, once again, like to thank all contributing authors, conference presenters, and referees. We would also like to thank Nick Wolterman and Charlotte Wilson at Emerald Publishing for their assistance and guidance. Finally, we are very grateful for financial support from Emerald Publishing, Eviews, Orion Construction, the International Association for Applied Econometrics, UC Irvine's Department of Economics, School of Social Sciences and Donald Bren School of Information and Computer Science, and the Purdue University Research Center in Economics.

Ivan Jeliazkov
Justin L. Tobias