Introduction

Subal C. Kumbhakar · Christopher F. Parmeter

The field of semi- and nonparametric econometrics is rapidly expanding. Similar to Moore's law, the number of papers offering new approaches using semi- and nonparametric econometrics has grown since the seminal seeds of these methods were discussed in the late 1950s and early 1960s. Figure 1 details the number of hits in Google Scholar for a simple search of the word `nonparametric` for each decade going back to the 1960s. As is clear, the growth of this term has been exponential. The procedures, tests, and methods being developed quickly outpace detailed references specifically designed to summarize this vast literature.

To adequately summarize this literature, numerous textbooks have been written (see Pagan and Ullah 1999; Li and Racine 2007; Henderson and Parmeter 2014) as well as monographs (Racine 2008) and collected works (Li and Racine 2009; Racine et al. 2014). However, given the pace at which new methods appear, an empirical *bergschrund* can arise, leaving practitioners with little guidance on appropriate implementation and use. This has the downside to render these methods empirically obsolete if not consistently distributed. Moreover, one cannot anticipate that each econometric breakthrough is followed by a textbook or monograph. Thus, it is crucial that advances in methods are critically applied, illustrating their use for a broad audience.

The current issue of *Empirical Economics* was commissioned to showcase implementation of cutting edge semi- and nonparametric methods across a wide array of applied economic domains. The allure of many of these papers is that they go beyond

S. C. Kumbhakar (✉)
Department of Economics, State University of New York, Binghamton University,
4400 Vestal Parkway East, Binghamton, NY 13902, USA
e-mail: kkar@binghamton.edu

C. F. Parmeter
University of Miami, Coral Gables, FL 33124-6520, USA
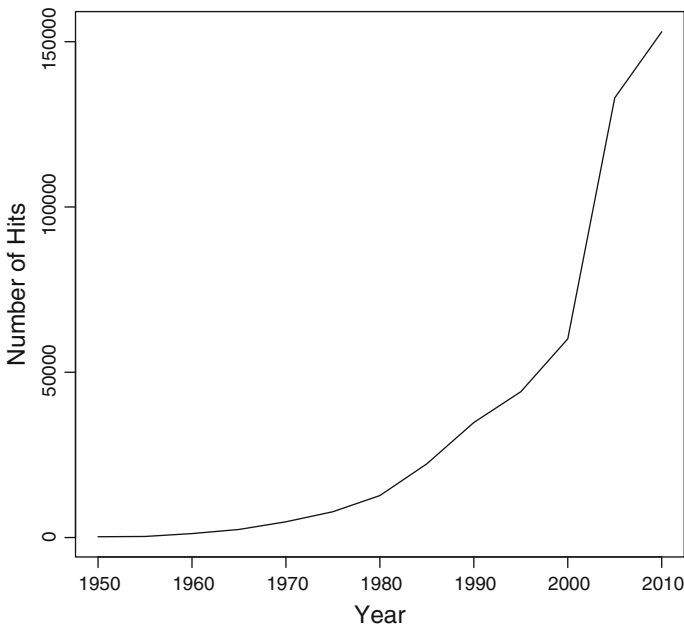e-mail: c.parmeter@miami.edu

Fig. 1 Google Scholar hits for key word `Nonparametric`

simple univariate density or regression estimation, which are perhaps the most commonly understood of the nonparametric tools available. The current collection of methods set forth on data is impressive. The contributors use copulas, additive regression, generalized polynomial regression, threshold modeling, data envelopment analysis, factor models, stochastic dominance, and more. We were fortunate to have many leading econometric scholars, providing studies across a wide array of economic domains, contribute to this special issue.

We were also fortunate to have an army of capable referees who provided excellent criticism and feedback on the research here, which undoubtedly improved the quality of the special issue. We would like to personally thank the referees for their dedicated service in helping us put this special issue together and ensuring the submissions were of sufficient quality. Without their expertise and assistance, this project would have been more arduous, and while our names appear on the special issue, the referees also deserve credit for this publication.

Given the depth and breadth of the applications here, we had a tough task to organize the papers. However, from our view, we chose to categorize the research into several distinct areas. First, we had several papers focusing on various aspects of banking as well as returns to education. Second, we have papers on copula estimation and the modeling of dependence. Third, there are several papers focusing on prediction and forecasting. Lastly, we have a wide array of papers that can aptly be connected as novel applications of cutting edge methodology.

Sainan Jin, Liangjun Su, and Yonghui Zhang propose a class of nonparametric tests of the existence of anomaly effects in empirical asset pricing models. They deploy

these tests within the framework of nonparametric panel data models with interactive fixed effects. This new class of tests is important as myriad of empirical findings suggest that the CAPM beta does not fully capture asset returns; these patterns in returns, which are not captured by theory, are known as anomaly effects. Their approach has two prominent features: the adoption of a nonparametric functional form to capture the anomaly effects and the flexible treatment of both observed/constructed and unobserved common factors. By estimating the unknown factors, betas, and nonparametric relationship simultaneously, their setup is robust to misspecification of functional form and common factors and avoids the common "error-in-variable" problem associated with the ubiquitous two-pass procedure. The approach finds that anomaly effects have a strong nonlinear presence across an array of factor models considered; this implies that commonly used linear empirical asset pricing models may be inappropriate.

Knowledge of the dependence structure within a distribution is paramount across many application domains, including health and finance. Under multivariate normality, computation of correlations is more than sufficient. However, deviations from normality call for more substantial methods. Herein lies value to semi/nonparametric approaches to discerning dependence. Jeffery Racine proposes a novel kernel-based estimator for the copula density that respects all of the key properties of a density (positivity, integration to 1). This new estimator is simple to estimate and analyze; a bevy of simulated and empirical examples are presented. Moreover, this new method is available in the np package (Hayfield and Racine 2008) in the freely available R programming language. Ximing Wu, Yichen Gao, and Yu Zhang also present a copula estimator, using a penalized exponential series-based approach to determine intergenerational dependence between parents and children body mass index. Their main finding is that the dependence relationship is asymmetric, suggesting persistent intergenerational transmission of being overweight or obese.

Food security and obesity are two significant public health issues; however, little is known about how these issues are connected. Daniel Millimet and Manan Roy use recently developed nonparametric bounds, in the context of causal modeling, to study the impact of food insecurity on childhood obesity. A benefit of the bounding approach is the ability to account for various forms of selection and reporting errors in the data. Their main finding is that while the bounds can sign the causal effect, accounting for estimation error, the confidence bounds never produce intervals that can sign the treatment effect. This suggests that food insecurity has an ambiguous effect on childhood obesity, which is an important insight for policy if even 1 % of households misreport their food security status.

More than three years after the Dodd–Frank Wall Street Reform and Consumer Protection Act was signed into law, regulators, policymakers, and academics in the USA, the UK, and elsewhere are still pondering the possibility and desirability of limiting the size of large banking organizations. Recent speeches by top Federal Reserve Bank officials in the USA and the Bank of England in the UK revived the unsettled debate on 'too big to fail' and the feasibility of limiting the scale and scope of bank activities, calling for further research on banking industry structure in general and on economies of scale and scope in particular.

Given this backdrop, Diego Restrepo-Tobón and Subal Kumbhakar use nonparametric regression methods to uncover returns to scale of US financial institutions.

They derive new measures of returns to scale based on input distance functions. In line with conventional wisdom, they find that not all bank holding companies and commercial banks enjoy increasing returns to scale (IRTS). In addition, economies of scale for those banking organizations operating under IRTS are small. Their RTS estimates are generally close to unity. Thus, despite the presence of IRTS for some of the biggest banking organizations, they conclude that the cost of breaking up some of these institutions into smaller and more manageable organizations may not impose heavy costs on the economy.

Pavlos Almanidis, Giannis Karagiannis, and Robin Sickles examined efficiency of large US commercial banks. Although efficiency studies in banking is not new, there is no consensus on how to model temporal behavior of inefficiency in panel models. They present a new model of time-varying inefficiency in which they modify Cornwell et al. (1990) specification to allow for switching patterns in temporal changes, using spline functions. The spline function specification can accommodate more than one turning point, thus allowing a non-monotonic relationship in technical inefficiency over time. They apply this model, using a translog input distance function, to a sample of large (too-big-to-fail) US commercial banks. Their empirical results reveal the presence of twelve break points over the 1984–2009 period, indicating a highly non-monotonic time pattern of average technical efficiency. They also find that the average bank in their sample has been operating at constant returns to scale since the mid 1990s.

On a much smaller banking scale, Michael Delgado, Christopher Parmeter, Valentina Hartarska, and Denis Nadolnyak use the generalized local polynomial methods of Hall and Racine (2014) in the context of a semiparametric smooth coefficient model to estimate economies of scope for a sample of microfinance institutions across the globe. They also present a unique approach to bandwidth selection in this context. Their main findings are that microfinance institutions display economies of scope, but many of these institutions do not offer both savings and loans. This offers these institutions options to expand their business model and extend their reach.

Returns to schooling can be defined as (a) the private return, (b) the social return, and (c) the labor productivity return. The main component in each of these measures is the impact of schooling on earnings where myriad statistical methods have been developed and fruitfully deployed. The most common specification in (a) includes log earnings on the left-hand side and schooling, experience and the square of experience on the right-hand side of the regression equation. A linear schooling variable suggests that an additional year of schooling gives the same percentage increase in earnings independent of the level of education. Many authors proposed alternative models to capture possible nonlinearities and heterogeneity.

Deniz Ozabaci and Daniel Henderson relax the functional form assumptions on both schooling and experience. Instead of deploying a fully nonparametric model, they used the nonparametric additively separable model with interaction terms and additional linear controls to look at changes in earnings with respect to increases in schooling. Their results confirm past research that heterogeneity in returns to schooling estimates exists both across and within standard subgroups (race, sex, and marital status). They also confirm past evidence that minorities receive higher rates of return to schooling on average than non-Hispanic whites. Differing from previous research, they find,

for married males, while non-Hispanic whites have lower returns on average, they typically possess the highest returns in the sample.

Feng Yao and Junsen Zhang consider estimating the effect of education on labor market earnings. In this context, an interesting puzzle is known (Card 2001) where the 2SLS estimate for the return to schooling typically exceeds the OLS estimate, but the 2SLS estimate is fairly imprecise. Their explanation is that it could be due to the restrictive linear functional form specification in the reduced form. They propose a kernel-based, semiparametric IV estimator for the parameters of the endogenous regressors to relax functional form restrictions. Applying the proposed estimators to estimate the return to schooling (as in Card 1995), they find that the semiparametric estimate of the return to schooling are much smaller and more precise than the 2SLS estimate, and the difference can largely be attributed to the misspecification in the linear reduced form.

High-performance computing instrumentation has become increasingly important in scientific research, and its use in research has expanded across diverse disciplines in recent years. Yet, even with the importance of high-performance computing, few papers have examined the impact it has on research productivity. To determine which academic departments benefited from their access to high-performance computing instrumentation, Amy Apon, Linh Ngo, Michael Payne, and Paul Wilson deploy new insights on nonparametric boundary estimators to empirically test if access to super computers leads to higher research productivity for a sample of American universities. The tests are applied across different departments, and their results find that not all departments benefit equally from access to high-performance computing. Specifically, chemistry, physics, and various engineering departments are more efficient in producing research where access to high-performance computing is available. Given the large sums of money that are allocated on an annual basis by both government agencies and academic institutions, these findings can shed light on appropriate allocations and directions for funding.

Despite its popularity among theoretical and experimental economists, the second price auction was more of a theoretical object until the appearance of eBay. The emergence of e-commerce through eBay and other online auction houses has prompted the attention of econometricians. Appropriate estimators for these auctions require care as there exist many aspects of the data that render standard regression methods mute. Wenchuan Liu, Yu Zhang, and Qi Li propose an econometric model of the price processes from second price online auctions. Given that auction bids monotonically increase within individual auctions but can differ considerably across auctions, they suggest using a monotone series estimator with a common relative price growth curve and auction-specific slopes. Furthermore, because the impacts of auction-specific attributes may evolve along the course of auctions, they employ a varying coefficient approach to accommodate these possibly time-varying effects. The proposed model is applied to eBay auctions of a Palm PDA. The estimates capture closely the overall pattern of online auction price processes, in particular, the bidding drought midway through auctions and the bid acceleration associated with bid sniping (last-minute bidding) at the end of auctions.

The analysis of the distribution of firms has a long history in economics. However, most of the existing distributional analysis of firms is static, and it is conceivable

that incorporating dynamics can lead to deeper insights. Kim P. Huynh, David Jacho-Chavez, Robert Petrunia, and Marcel Voia present a functional principal component estimator to study distribution dynamics for Canadian firms. They determine that firm debt-to-asset ratio distributions have persistent deviations from the initial distributions, and bootstrap-based tests confirm that these distributions are statistically different. Their method illustrates the efficacy of functional principal components in applied research and specifically, to the study of the distribution of firms across the economic landscape.

Isabel Proença, Stefan Sperlich, and Duygu Savaşci present a semiparametric estimator for the ubiquitous gravity model of international trade. Their proposed estimator is an alternative to the pseudo-Poisson maximum likelihood estimator of Santos Silva and Tenreyro (2006). In the spirit of the correlated random effects linear regression estimator, their gravity model estimator marks a compromise between the common fixed effect and random effect frameworks. This estimator is easily deployed using standard penalized spline methods and may prove invaluable in short panels for trade data where there is limited time variation.

Testing the predictive ability of a given forecasting procedure against a group of alternative forecasting procedures is important in applied time series analysis. Zongwu Cai, Jiancheng Jiang, Jingshuang Zhang, and Xibin Zhang propose a new test for superior predictive ability based on the seminal work of Fan and Jiang (2007) on semiparametric generalized likelihood ratio tests. Their test has improved power over existing tests of data-snooping bias. Further, the test is applicable to a variety of multiple testing scenarios, such as the performance of mutual fund managers. Alternatively, Nan Cai, Zongwu Cai, Ying Fang, and Qiuhua Xu propose a new semiparametric smooth transition autoregressive model by allowing state variables to enter into the transition function in a completely nonparametric fashion. Their empirical results, based on forecasting five major Asian exchange rates, show that the new model has some advantages in out-of-sample forecasting performance.

A large body of empirical studies have examined the statistical properties of the real exchange rate to draw implications on the practical relevance of purchasing power parity (PPP) over the long-run. And while traditional linear cointegration tests of the PPP hypothesis often lead to rejection, more recent studies allowing for nonlinearities suggest mixed results, leaving the outcome of the hypothesis unresolved. Using a semiparametric varying coefficient framework, Hongjun Li, Zhongjian Lin, and Cheng Hsiao apply recently developed cointegration tests to study the PPP hypothesis between the USA and Canada, Japan, and the UK. Contrary to typical findings based on linear models, the semiparametric model provides support for the PPP hypothesis across all three country comparisons with the USA.

Typical measures of wealth focus on capital stocks, health, and institutional quality. However, these notions of wealth do not capture sustainability, ignoring changes in stocks of resources over time and the health of the environment. Elettra Agliardi, Mehmet Pinar, and Thanasis Stengos employ stochastic dominance methods to derive a relative environmental degradation index across countries. This index captures one of the eight dimensions suggested by Stiglitz et al. (2009) that go beyond gross domestic product to measure well-being. A worst-case scenario index to measure environmental degradation across different countries at different times is constructed using methods

based on consistent tests for stochastic dominance efficiency. It is found that in the worst-case scenario index, green house gas emissions contribute the most and water pollution the least. Their relative worst-case scenario index can be a useful tool for policy making in conveying information on environmental quality and enables quick assessment of sustainable performance across countries and over time. Further, this index can serve as an important benchmark to assess the progress that countries make in reducing their environmental risk.

The adverse macroeconomic effects of volatility on economic growth are well known. However, a variety of empirical findings fail to reach consensus on the most crucial sources of growth volatility, potentially due to theory and model uncertainty. Chih-Ming Tan and Andros Kourtellos present a Bayesian model averaging estimator for growth volatility in the presence of a threshold. Their approach develops a novel empirical methodology that allows the simultaneous handling of both theory uncertainty and parameter heterogeneity. They find evidence of both parameter heterogeneity and model uncertainty, highlighting the role of ethnic fractionalization, institutions, financial development, health, and geography on volatility of cross-country growth.

Geraldine Henningsen, Arne Henningsen, and Christian Henning present a model of transaction costs and social networks among firms. It is argued that firm productivity is downward bias if these underlying, hidden social costs are not accounted for. They present estimates from a range of semi- and nonparametric models to discern just how large this bias is. They also determine that large trading networks and dense household networks have a positive influence on a farmer's productivity. Furthermore, transaction costs are determined to have a measurable impact on the productivity ranking of the farms.

As a final note, we would like to thank Liane Wolf for behind-the-scenes assistance at various stages as well as members of the editorial board for approving this project. Without their belief, this project would not have come to fruition.

References

Card D (1995) Using geographic variation in college proximity to estimate the return to schooling. In: Christophides LN, Grant EK, Swidinsky R (eds) Aspects of labour market behavior: essays in honour of John Vanderkamp. University of Toronto Press, Toronto, pp 201–222

Card D (2001) Estimating the return to schooling: progress on some persistent econometric problems. Econometrica 69:1127–1160

Cornwell C, Schmidt P, Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. J Econ 46(2):185–200

Fan J, Jiang J (2007) Nonparametric inference with generalized likelihood ratio tests (with comments). Test 16:409–444

Hall P, Racine JS (2014) Infinite-order cross-validated local polynomial regression. J Econ. (forthcoming)

Hayfield T, Racine JS (2008) Nonparametric econometrics: The np package. J Stat Softw 27(5). http://www.jstatsoft.org/v27/i05/

Henderson DJ, Parmeter CF (2014) Applied nonparametric econometrics. Cambridge University Press, New York

Li Q, Racine J (2007) Nonparametric econometrics: theory and practice. Princeton University Press, Princeton

Li Q, Racine JS (2009) Advances in econometrics: nonparametric econometric methods, vol 25. Emerald Insight, New York

Pagan A, Ullah A (1999) Nonparametric econometrics. Cambridge University Press, New York

Racine JS (2008) Nonparametric econometrics: a primer. Found Trends Econ 3(1):1–88
Racine JS, Su L, Ullah A (2014) Oxford handbook of semiparametric and nonparametric econometric methods. Oxford University Press, New York
Santos Silva J, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88(4):641–658
Stiglitz JE, Sen A, Fitoussi J-P (2009) Report by the commission on the measurement of economic performance and social progress. http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf