ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Editorial

Introduction to the Annals Issue in Honor of Gary Chamberlain

This Annals Issue honors Professor Gary Chamberlain, who was the Louis Berkman Professor of Economics at Harvard University. Gary was a profound and deep thinker who made foundational contributions to latent variable analysis, panel data, quantile regression, semiparametric efficiency theory, and statistical decision theory in econometric problems and asset pricing among other areas. In addition to being an outstanding scholar, Gary was a dedicated teacher and mentor.

In 2018, a conference in Gary's honor was held at Harvard University. The conference provided an opportunity for Gary's collaborators, colleagues and students to pay tribute to him. Several of the papers included in this special issue were originally presented at this conference.

This issue begins with a contribution from Gary himself. "Feedback in Panel Data Models" first appeared as a Harvard Institute of Economic Research Working Paper (No. 1656) in 1993. This paper only circulated in hard copy form and, consequently, copies have been difficult to come by. Nevertheless the paper is an underground classic of sorts, having been cited by many leading surveys and key references in the dynamic panel data literature.

The paper connects two ideas developed by Gary in other work in the early 1990s. The first involves sequential moment conditions in panel data (Chamberlain, 1992). Sequential moment restrictions arise naturally in economic models where agents' information sets grow over time. For example, while firms may make current period input decisions prior to the realization of any innovation in total factor productivity (TFP), they generally condition on past TFP shocks when doing so. If TFP is persistent, then a firm's input choices will be (mean) independent of current and future innovations to TFP, but not past ones. This informational structure features prominently, for example, in modern empirical production function research (Arellano and Bond, 1991; Olley and Pakes, 1996; Blundell and Bond, 2000, e.g.). It is also relevant, as noted by Gary in his paper, in certain program evaluation settings.

The second idea relates to Gary's work on identification and efficiency of conditional moment restriction models with nonparametric components (Chamberlain, 1992). Gary provides positive results – identification with non-zero information bounds – for panel data models with increasing information sets over time and a *scalar* nonparametric component (corresponding to the conditional mean of a unit-specific intercept in his main example). For models with vectors of nonparametric components, as might arise in random coefficient models, he shows that the parameters are not identified in general.

This is an elegant paper and, while almost 30 years old, is brimming with interesting ideas that will no doubt inspire further research. We are delighted to publish this paper. Gary submitted the paper to this special issue prior to his passing. Because we were unable to locate Gary's original source computer files, the paper was re-typed in LaTeX using the pdf file Gary submitted to the journal. We are grateful to Elie Tamer and, especially Chris Walker, for help in re-typing the paper. A small number of typographical corrections aside, the version of the paper published here corresponds to the original. We are also grateful to Laura Gehl, Gary's daughter, for her encouragement and openness to making her father's unpublished research more widely available.

The other papers in this special issue are presented in alphabetical order by first author. In reflection of Gary's wide-ranging contributions to econometrics and empirical economics, these papers cover a variety of topics and employ different analytical and computational techniques. Many of them are grounded in applications across a number of areas in economics, finance, and health.

Panel data were a major focus of Gary Chamberlain's research throughout his career. In "Robust Likelihood Estimation of Dynamic Panel Data Models", Javier Alvarez and Manuel Arellano consider estimation of dynamic panel data models with an autoregressive structure and individual effects, which are important for analyzing persistence and heterogeneity in longitudinal data. They develop a random effects maximum likelihood approach that can handle heteroskedasticity over time in the model disturbances and does not require mean stationarity. They consider variations on the approach that impose different restrictions, compare their large sample variances, and develop an empirical application to individual earnings data from the European Community Household Panel.

https://doi.org/10.1016/j.jeconom.2021.09.001

Please cite this article as: B. Graham and K. Hirano, Introduction to the Annals Issue in Honor of Gary Chamberlain. Journal of Econometrics (2021), https://doi.org/10.1016/j.jeconom.2021.09.001.

In "Design-based Analysis in Differences-in-Differences Settings with Staggered Adoption", Susan Athey and Guido Imbens study designed-based inference for average treatment effects in a panel data setting. They focus on the case where units opt in to treatment at different times, but once treated remain treated thereafter. The simple difference-in-differences estimators considered in Section 4 of the paper are, of course, special cases of linear panel data estimators in Gary's *Handbook of Econometrics* chapter (Chamberlain, 1984) and many other papers in the literature. Athey and Imbens show that these methods estimate weighted averages of treatment effects in the context of a more general data generating process. This follows under the maintained assumption that the timing of treatment is random. They further develop design-based methods of inference, where statistical uncertainty is due to the randomness of treatment take-up date. This differs from the type of sampling-based approaches to inference featured in Chamberlain (1984). Interestingly, the "randomization variance" of Athey and Imbens generally results in smaller standard errors relative to those produced by cluster-robust variance estimators.

In their contribution "Estimation and Inference of Semiparametric Models using Data from Several Sources", Moshe Buchinsky, Fanghua Li, and Zhipeng Liao introduce minimum distance estimators that combine information from multiple datasets. For example, to study household consumption dynamics a researcher might combine consumption data from the Consumer Expenditure Survey (CEX) with wealth and asset data from the Health and Retirement Study (HRS). Data combination problems like these have been studied before. An example familiar to many economists is the two-sample instrumental variables (TSIV) estimator (see Graham et al. (2016) for more examples and references to the literature). Buchinsky, Li, and Liao extend prior work to accommodate conditional moment restrictions (and hence an infinite number of unconditional restrictions). Here they build on results by Ai and Chen (2003) and others. They further provide large sample theory and a consistent asymptotic variance estimator.

Survey data often suffer from missing data due to nonresponse. In "Minimax-Regret Sample Design in Anticipation of Missing Data, with Application to Panel Data", Jeff Dominitz and Charles Manski consider the problem of sample design from a decision-theoretic perspective. Specifically, they examine the choice of how to collect a two-period panel data set, where there is complete response in the first period, but there is nonresponse (attrition) in the second period. The data will be used to predict some function of the variables in both periods, under squared error loss; due to the nonresponse, the target value is not point-identified in general. The survey designer can choose between different sampling designs, for example between a design with low cost (enabling larger samples) but higher attrition, versus a higher-cost design with lower attrition.

Dominitz and Manski analyze this problem within the Wald statistical decision theory framework. They work with the minmax regret criterion, which evaluates decision rules by their maximum regret over the statistical model space. Focusing on linear and indicator functions of the variables, they develop results on the maximum regret for midpoint predictors of partially identified quantities, which enables comparison of alternative sampling schemes. The authors also examine the related problem of choosing between a panel survey design and a repeated cross-section design. They show that, depending on the type of function of the variables being predicted, a repeated cross section design can have lower or higher maximum regret than a panel design with equivalent sample size and cost.

A common goal underlying many econometric methods is to estimate the effect of some variable X on another variable Y, holding fixed other variables W. This idea appears in causal inference using the propensity score under the assumption of unconfoundedness when X is binary, and in the concept of an average partial effect for general X. In "Semiparametrically Efficient Estimation of the Average Linear Regression Function", Bryan Graham and Christine Campos de Xavier Pinto consider the best linear predictor of Y given X conditional on W, and seek to estimate the average of the linear predictor coefficients over W. Here X can be continuous or multidimensional, and the joint distribution of the variables is allowed to be flexible, but by focusing on the average linear predictor coefficient the problem becomes semiparametric. Graham and Pinto calculate the semiparametric efficiency bound for estimation of this quantity, and propose a new estimator that is locally efficient. Their estimator is simple to compute, improves upon some prior proposed estimators, and generalizes familiar Oaxaca–Blinder type covariate adjustment methods.

In "Analyzing Cross-Validation for Forecasting with Structural Instability", Keisuke Hirano and Jonathan Wright study the use of cross-validation for determining how much time series data should be used to forecast future observations. Limiting the amount of past data used can mitigate bias in the presence of parameter instability, but it also increases variance. The paper studies this trade-off. Specifically they examine the risk properties of forecasts made when cross validation is used to select the data window. Gary had a long standing interest in empirical risk minimization procedures. Indeed he talks about working on a basic problem of this type as an undergraduate research assistant for Martin Feldstein in his *Econometric Theory* Interview (Graham et al., 2021).

Maximilian Kasy's paper, "Who Wins, Who Loses? Identification of the Welfare Impact of Changing Wages", studies how to measure the welfare effects of policies changes when behavioral responses may be present. His motivating (and empirical) example involves assessing the effects of an expansion of the Earned Income Tax Credit (EITC) in the United States. This setting requires (i) taking into account any behavioral response by EITC-qualifying workers (presumeably an increase in their labor supply), (ii) the decline in market wages due to any increase in labor supply and, finally, (iii) any welfare impacts on non-EITC-qualifying workers who compete in the same labor market as EITC-qualifying ones. The paper presents conditions under which vectors of "outcome-conditioned" causal effects are identifiable. To achieve point identification, a restriction on the nature of heterogeneity across agents is required (see Theorem 3). Kasy's work extends prior work by Hoderlein and Mammen (2007) and, importantly, shows how it can be used to undertake welfare analysis.

Statistical decision theory provides a powerful set of tools for analyzing statistical problems, but applying it directly to empirical problems is not always feasible due to computational limitations. (See Chamberlain (2000) for an example of computational methods for statistical decision theory.) In "Approximation of Sign-regular Kernels", Thomas Knox considers the problem of evaluating parametrized integrals of the form $\int K(x, y)dP(x)$, which arise in a range of applications, including option pricing. He considers approximations to such integrals of the form $\sum_i c_i g_i(y)$. This form is of lower rank and can reduce the computational burden, provided that the terms in the sum can be chosen to give small approximation error. For economic applications involving optimization, it is desirable to minimize *uniform* approximation error. With this motivation, Knox shows how to numerically construct approximations that, under certain conditions on the kernel being approximated, minimize the uniform approximation error. This approach leads to much lower uniform error than existing methods that target mean-squared or local approximation error, as illustrated in applications to portfolio choice and prior elicitation. An open-source software package for this algorithm, *KernelyzeBase*, is available.

Quantile regression (Koenker and Bassett, 1978) provides a rich, flexible framework for empirical work, which Gary recognized and explored in Chamberlain (1994). In "Censored Quantile Regression Survival Models with a Cure Proportion", Naveen Narisetty and Roger Koenker consider survival time models where some fraction of the population may become "cured" with an effectively infinite survival time. They model the survival time (or a transform of the survival time) as a quantile regression conditional on covariates. There may be a positive conditional probability of a cure, and the survival times can have random censoring.

Narisetty and Koenker propose a data augmentation algorithm to estimate the model. In one step of the algorithm, the latent indicator for being cured is simulated conditional on the current values of the model parameters. The quantile regression model for survival times involves a continuum of parameters (one vector for each quantile level); the algorithm handles this by discretizing the model over a grid of possible quantile levels. This estimation method is implemented in an R package, *quantreg*. Model identification is discussed, and the paper includes a simulation study and an application to lung cancer data.

Assembling this special issue was a group effort. We wish to thank the contributors to the issue, who patiently endured a rather long editorial process. We also want to acknowledge the work of many gracious referees, the co-editors and editorial staff of the *Journal of Econometrics*, and Laura Gehl. Finally, on behalf of his former students, colleagues and co-authors, we would like to thank Gary for many inspiring conversations. He is sorely missed by all of us.

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843.
Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econom. Stud. 58 (2), 277–297.
Blundell, R., Bond, S., 2000. GMM estimation with persistent panel data: an application to production functions. Econometric Rev. 19 (3), 321–340.
Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. 2. North-Holland, Amsterdam, pp. 1247–1318.
Chamberlain, G., 1992. Comment: Sequential moment restrictions in panel data. J. Bus. Econom. Statist. 10 (2), 20–26.
Chamberlain, G., 1994. Quantile regression, censoring, and the structure of wages. In: Sims, C. (Ed.), Advances in Econometrics: Sixth World Congress, vol. I. Cambridge University Press, Cambridge, pp. 171–209.
Chamberlain, G., 2000. Econometric applications of maxmin expected utility. J. Appl. Econometrics 15 (6), 625–644.
Graham, B.S., Hirano, K., Imbens, G.W., 2021. ET interview: Professor Gary Chamberlain. Econometric Theory (forthcoming).
Graham, B.S., Pinto, C., Egel, D., 2016. Efficient estimation of data combination models by the method of auxiliary-to-study tilting (AST). J. Bus. Econom. Statist. 31 (2), 288–301.
Hoderlein, S., Mammen, E., 2007. Identification of marginal effects in nonseparable models without monotonicity. Econometrica 75 (5), 1513–1518.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.
Olley, S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263–1297.

Bryan Graham
Department of Economics, University of California-Berkeley, 530 Evans Hall, #3380, Berkeley, CA 94720-3888, United States of America
National Bureau of Economic Research, United States of America
E-mail address: bgraham@econ.berkeley.edu.

Keisuke Hirano *
Department of Economics, Pennsylvania State University, 602 Kern Graduate Building, University Park, PA 16802, United States of America
E-mail address: kuh237@psu.edu.

Available online xxxx

* Corresponding editor.