Innovations in Measurement in Economics and Econometrics: An Overview

Co-Editors:

W. A. Barnett, W. E. Diewert, and E. Maasoumi

Annals Issue of the *Journal of Econometrics*

The papers in this Special Annals Issue represent a remarkable collection of contributions by leading scholars in their respective areas. The range of topics covered is as broad as the numerous conceptual and methodological issues that are expertly addressed by these authors. The common theme of "measurement in economics," and innovative development and insights is well served and demonstrated.

Many of the authors are the leading proponents of the topical areas addressed, and offer state of art econometric and statistical treatments of a number of significant economic measurement and decision issues, including several aspects of aggregation and index numbers, measurement problems in GDP and other leading indicators, multiple indicator/latent object problems, identification in nonparametric settings, examination of treatment effects and dynamic continuing values, classification in poverty measurement and uniform ranking of outcomes. The contributions are very much in the spirit of the ideals espoused by the new Society for Economic Measurement (SEM).

The papers appearing in this important Annals Issue are those that managed to successfully pass through rigorous refereeing that is routine to *Journal of Econometrics* and its highest standards. We are grateful to a large number of referees and authors for participating in this fruitful process.

In the paper, "Dynamic Treatment Effects," James J. Heckman, John Eric Humphries, and Gregory Veramendi develop robust models for estimating and interpreting treatment effects, in ordered and unordered multistage decision problems. Identification strategies are based on instrumental variables and/or conditional "matching" assumptions. Importantly, treatment effects are decomposed into direct effects and continuation values associated with moving to the next stage of a dynamic decision problem. This is novel and consequential. In this framework, the authors decompose the IV estimator, and show that IV generally does not estimate economically interpretable or policy relevant parameters in prototypical dynamic discrete choice models, unless policy variables are instruments. This too has major implications for understanding of the findings in empirical treatment effect literature. For instance, "continuation values" are an empirically important component of estimated total treatment effects of education. The authors employ their methods to estimate the components of what LATE estimates in a dynamic discrete choice model.

Charles F. Manski's "Credible Interval Estimates for Official StatisticsWith Survey Nonresponse" argues that government agencies commonly report official statistics based

on survey data as point estimates, without accompanying measures of error. It is more challenging for agencies to measure "nonsampling" errors, whereas they could measure sampling error using established statistical principles. His paper considers error due to survey nonresponse. The standard practice has been to use weights and imputations to implement assumptions that nonresponse is conditionally random. Manski reviews modern research deriving *interval* estimates without assumptions about the values of missing data. He uses data from the U. S. Current Population Survey to form interval estimates for median household income, the family poverty rate, and the unemployment rate. These estimates allow an examination of the middle ground between his interval estimation and point estimation assuming that nonresponse is conditionally random. Agencies concerned with official statistics and measures can benefit by reporting analogous measurements.

In "On Independence Conditions In Nonseparable Models: Observable and Unobservable Instruments," Rosa L. Matzkin develops identification results employing independence conditions among un-observable variables. The independence conditions are used to derive first-stage nonseparable reduced form functions. These functions are employed to express the derivatives of nonseparable structural functions in terms of the derivatives of the reduced form functions. When endogeneity is due to simultaneity, Matzkin provides new results by combining the independence assumptions with parametric specifications and exclusion restrictions. If there is also triangularity, all functions can be nonparametric and nonseparable in unobservable random terms. For the latter situation, Matzkin offers several equivalence results and illuminates the trade-offs between observable and unobservable instruments.

This paper addresses a currently active area of research that is illuminating the information content of various restrictive assumptions that are required for inference in otherwise challenging, nonseparable functional settings that are actively examined for examination of labor market outcomes and auctions, among others.

In the paper "Real-Time Nowcasting of Nominal GDP with Structural Breaks," William A. Barnett, Marcelle Chauvet, and Danilo Leiva-Leon provide early assessments, nowcasts, of current U.S. Nominal GDP growth. This is a potential new monetary policy target. Nowcasts have the advantage of being computed based on the exact amount of information that policy makers have available at the time predictions are made. Since real time information arrives at different frequencies and is asynchronous, one faces the challenge of mixed frequencies, missing data, and ragged edges. This paper proposes a multivariate state space model that accommodates asynchronous information as well as parameter instability (DYMIBREAK). The authors employ small scale confirmatory factor analysis to select candidate variables based on their ability to forecast nominal GDP. The model is fully estimated in one step using a nonlinear Kalman filter. In contrast to principal component analysis, the factor model captures comovements rather than the variance underlying the variables. The predictive ability of the model is compared with other univariate and multivariate specifications. They find that the proposed model, containing information on real economic activity, inflation, interest rates, and Divisia monetary aggregates, produces the most accurate real time nowcasts of nominal GDP growth.

Another paper considering asynchronous data and measurement errors is “Estimating the Quadratic Covariation Matrix for Asynchronously observed high frequency stock returns corrupted by Additive Measurement Error” by Sujin Park, Seok Young. Hong, and Oliver Linton. Their contribution is to study the estimation problem of the covariance matrices of asset returns in the presence of microstructure noise and asynchronicity between the observations across different assets. They propose a new Fourier domain based estimator of multivariate ex-post volatility, called the Fourier Realized Kernel (FRK). An advantage of this approach is that no explicit time alignment is required unlike the widely used time domain methods. The authors derive large sample properties and establish asymptotic normality of their estimator under some general conditions that allow for both temporal and cross-sectional correlations in the measurement error process. This is a Frequency domain extension of the asymptotics for the multivariate realized kernel estimator in Barndorff-Nielsen et al. (2011). Based on simulations, it is shown the method outperforms the time domain estimators when two assets with different liquidity are traded asynchronously.

In “A new Approach to measuring and studying the characteristics of class membership: Examining poverty, inequality and polarization in Urban China,” Gordon Anderson, Alessio Farcomeni, Maria Grazia Pittau and Roberto Zelli embark on classifying agents into subgroups in order to measure the plight of the “poor,” “middle class,” or “rich”. The definition of class boundaries is contentious and beset with problems. The authors propose a technique based on mixture models to overcome these problems by determining the number of classes in a population and estimating the probability that an agent belongs to a particular class. The possibility of studying the correlates of class membership is examined as well as other desirable statistics. Household income in Urban China in the last decade of the 20th Century is examined. Four income groups are classified and the progress of those “poor,” “lower middle,” “upper middle,” and “rich” classes are related to household and regional characteristics. This analysis is state of art and rigorous.

Another contribution to the pressing topic of poverty and relative standings is “Consistent Tests for Poverty Dominance Relations,” by Garry F. Barrett, Stephen G. Donald, and Yu-Chin Hsu. Their paper considers methods for comparing poverty in two income distributions. A discussion of the concept and usefulness of the Poverty Gap Profile (PGP) makes clear that dominance of one PGP over another suggests poverty dominance for a wide class of indices which may be expressed as functionals of the PGP. This allows casting hypotheses on poverty dominance in terms of the PGP and introduce a test based on empirical PGPs. The poverty line is in fact estimated, and a method for obtaining critical values by simulation is proposed that takes account of estimation of the poverty line. The finite sample properties of the methods are presented by Monte Carlo simulation. The paper presents an examination of relative consumption poverty in Australia over the period 1988/89-2009/10.

The paper, “A Solution to Aggregation and an Application to Multidimensional ‘Well-Being’ Frontiers,” by Esfandiar Maasoumi and Jeffrey S. Racine, proposes a new technique for identification and estimation of aggregator functions in multidimensional evaluations and multiple indicator settings. In most applied contexts, objectives such as

well-being or performance are multidimensional. Choosing weights for them has been a long standing problem. Subjectively chosen weights are commonly used. That procedure is far from optimal. This paper provides a new and much superior approach. Technical advances allow nonparametric inference on the joint distribution of continuous and discrete indicators of well-being. In a multi-attribute setting, quantiles are frontiers that define equivalent sets of covariate values. The authors identify these frontiers nonparametrically at first. Then they suggest parametrically equivalent characterizations of these frontiers that reveal suitable weights for different attributes for different groups and different quantiles. These estimated parametric functions are "ideal aggregators" in a sense defined by the authors.

The paper proposes parametric estimation, using the CES form of the ideal aggregator function, of the non-parametrically estimated level sets. Constant elasticity of substitution provides the implicit prices of attributes for the level set. This approach thereby provides measures of how the different attributes contribute to welfare at the chosen welfare level. To compare different groups of sub-populations, it is possible to carry out multivariate stochastic dominance tests. An illustration is provided with Indonesian data analyzing multidimensional poverty. This new approach resolves a classic problem of assigning weights to multiple indicators, as well as empirically incorporating the key component into multidimensional analysis.

S. Boragan Aruoba, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, and Dongho Song provide a new measure of U. S. GDP growth in their paper, "Improving GDP Measurement: A Measurement-Error Perspective." Their measure uses both expenditure-side and income-side GDP estimates. Two competing measures of output exist: GDP and GDI. They are numerically different. This paper uses filtering methods allowing for measurement error to extract a latent output measure. This may be interpreted as another example of multiple indicators for a latent concept, as in the Maasoumi-Racine paper above. The blended approach applies optimal signal-extraction techniques to the noisy expenditure-side and income-side GDP estimates. The new measure often differs substantially from the official measures. The dynamic properties of the new measure also differs noticeably from those of the traditional measures. In particular, the persistence of the aggregate output dynamics is stronger than previously thought.

The final four papers are concerned with various aspects of index number theory; i.e., exactly how should price and quantity aggregates of consumption, output, or input be constructed? *Bilateral index number theory* addresses how price and quantity aggregates can be constructed for comparisons between two situations, where the detailed price and quantity data could refer to consumption or production at two different points in time for the same group of economic agents, or the data could refer to comparisons across two regions. *Multilateral index number theory* addresses how price and quantity aggregates could be constructed across many time periods or many regions. It turns out that all four papers draw on multilateral index number theory.

The paper by Kevin J. Fox and Iqbal A. Syed, "Price Discounts and the Measurement of Inflation" introduces a new potential source of bias in the construction of a Consumer Price Index (CPI), namely the *exclusion of sales prices* when constructing the CPI. Many countries do not collect the prices of commodities sold at discounted prices in retail outlets when they collect samples of prices to construct their CPIs. Obviously, the omission of sales prices will lead to CPIs with price *levels* which are too high, but it is not clear if this omission will lead to upward biases in CPI *inflation* rates, since the upward bias in the price levels could remain roughly constant from period to period. However, Fox and Syed found that ignoring sale prices led to an annual upward bias in their overall chained Törnqvist index of 0.56 percentage points per year and an upward bias in their overall Rolling Year index of 0.61 percentage points per year.

The paper by Robert J. Hill, "A Least Squares Approach to Imposing Within-Region Fixity in the International Comparisons Program," addresses an interesting practical problem in multilateral index number theory. The World Bank's International Comparisons Program (ICP) undertakes periodic comparisons of prices (in constant units of measurement) of the components of final demand GDP for most countries in the world. For the last two rounds of the ICP, the World Bank breaks the world up into 6 regions. Each of the 6 regions constructs its own comparison of prices across the countries in its region. The problem Hill addresses is this: how exactly should the regions be linked up so that the real GDP of all countries can be compared but at the same time, the relative real GDPs of countries within a region are preserved? Hill calls this the *fixity problem* and he develops a new method (with attractive properties) for solving this problem.

The third paper on index number theory, "Stochastic Approach to Computation of Purchasing Power Parities in the International Comparison Program (ICP)," is by D.S. Prasada Rao and Gholamreza Hajargasht. These authors substantially advance the stochastic approach to index number, and they use the method of moments (MOM) to derive the Geary-Khamis, Iklé, Rao and other multilateral index number methods that are commonly used in international comparisons. Expressions for computing standard errors for the country price indexes based on these formulae are also derived.

The fourth paper on index number theory, "Measuring Industry Convergence and Cross-Country Convergence," is by Robert Inklaar and W. Erwin Diewert. This paper compares multifactor productivity levels for 38 countries over the time period 1995-2011 for the market sector of each economy and two subsectors of the market sector. The basic questions addressed in this paper are as follows: (i) are countries converging to the world productivity frontier over this time period and (ii) is the convergence different in the traded and nontraded sectors? Their paper also addresses a problem with the standard method for comparing productivity levels across countries. This method uses distance functions to aggregate outputs and inputs, and thus it cannot be used to compare levels of GDP across countries, since distance functions are not well defined when the output aggregate has a negative output component (like imports).

We thank Cheng Hsiao and the editorial Board of *Journal of Econometrics* for their support and encouragement.