



## Introduction

Subal C. Kumbhakar<sup>1</sup> · Robin Sickles<sup>2</sup> · Hung-Jen Wang<sup>3</sup>

Published online: 16 June 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

This Special Issue in Honor of Peter Schmidt was put in motion by Subal Kumbhakar who as a co-editor of *Empirical Economics* enlisted Hung-Jen Wang and Robin Sickles to join him as co-editors of the Special Issue. When we sent requests for submissions, based in large part on Peter's suggestions, the responses were quick and overwhelming. From the VERY best scholars in econometrics came comments that would soften the heart of the most cynical academics. These scholars grew their careers and accomplished such remarkable professional achievements while admiring and benefiting from Peter's writings, insights, and friendships. We are so very lucky to have been able to put together this special of *Empirical Economics* in Honor of Peter Schmidt and we thank the editors and Springer Nature for making this possible.

In our call for papers to selective potential submitters in late 2021, we pointed out that Peter Schmidt stepped down as an Associate Editor of *Empirical Economics* after serving in that position for over 24 years. We noted his distinguished service at the *Journal* and that among his lifelong accomplishments in academics were important contributions to many areas of econometric research, including time series econometrics, panel data econometrics, and stochastic frontier analysis. We also noted that his research always had a fine balance between theoretical innovation and empirical relevance, which is what is valued at *Empirical Economics*.

This Special Issue contains contributions on state-of-the-art topics in econometrics related to the three broad research areas mentioned above as well as a few others that, based on submissions, we felt deserved their own topic, namely his contributions to the literature on applied econometrics, copulas, and nonparametric methods, as well as on limited dependent variables.

---

✉ Subal C. Kumbhakar  
kkar@binghamton.edu

Robin Sickles  
rsickles@rice.edu

Hung-Jen Wang  
wangh@ntu.edu.tw

<sup>1</sup> Department of Economics, Binghamton University - SUNY, Binghamton, NY, USA

<sup>2</sup> Department of Economics, Rice University, Houston, TX, USA

<sup>3</sup> Department of Economics, National Taiwan University, Taipei 10617, Taiwan

The Special Issue is composed of 25 manuscripts. We will briefly discuss each paper in order of the number of contributions on the topics, which give an order to our summary that begins with panel data, then stochastic frontiers and efficiency/productivity measurement, time series, applied econometrics, copulas, and nonparametric methods, and finishes with limited dependent variables.

## 1 Panel data

In “Robust dynamic space–time panel data models using  $\varepsilon$ -contamination: An application to crop yields and climate change,” Baltagi, Bresson, Chaturvedi, and Lacroix extend the Baltagi et al. (2018, 2021) static and dynamic  $\varepsilon$ -contamination papers to dynamic space–time models. They investigate the robustness of Bayesian panel data models to possible misspecification of the prior distribution. Using an extensive Monte Carlo simulation study, they compare the finite sample properties of their proposed estimator to those of standard classical estimators. They obtain short-run as well as long-run effects of climate change on corn producers in the USA.

In “Unbiased estimation of the OLS covariance matrix when the errors are clustered,” Boot, Niccodemi, and Wansbeek derive an estimator that is unbiased when the random-effects model holds. They do the same for two more general structures and study the usefulness of these estimators against others by simulation, the size of the  $t$  test being the criterion.

In “Refined GMM estimators for simultaneous equations models with network interactions,” Egger and Prucha propose a refinement of the generalized spatial two-stage and three-stage least squares estimators for simultaneous systems of equations with network interdependence, introduced in Drukker et al. (2022). The refinement proposed involves the weighting of the moment conditions underlying those estimators.

In “Identification and estimation of categorical random coefficient models,” Gao and Pesaran propose a linear categorical random coefficient model, in which the random coefficients follow parametric categorical distributions. The distributional parameters are identified based on a linear recurrence structure of moments of the random coefficients. A generalized method of moments estimation procedure is proposed to address heterogeneity in time effects in panel data models (Ahn and Schmidt 1995; Ahn et al. 2013). The utility of the proposed estimator is illustrated by estimating the distribution of returns to education in the USA by gender and educational levels.

Han and Kim, in “Dynamic panel GMM estimators with improved finite sample properties using parametric restrictions for dimension reduction,” propose reducing finite sample bias by imposing parametric restrictions on the expected first derivative matrix and the covariance matrix of the sample moment functions. The resulting estimator is consistent under regularity irrespective of the correctness of the extra restrictions and is first-order efficient if they are indeed correct. The method is applied to a dynamic cigarette consumption model.

In “Testing for correlation between the regressors and factor loadings in heterogeneous panels with interactive effects,” Kapetanios, Serlenga, and Shin address whether the regressors are correlated with factor loadings or not. They explore this issue and

propose a Hausman-type test to address the matter. Further, they develop two nonparametric variance estimators for the FE and PC estimators as well as their difference that are robust to the presence of heteroscedasticity, autocorrelation, and slope heterogeneity.

Li, Li, and Hsiao, in their contribution “Assessing the impacts of pandemic and the increase of minimum down payment rate on Shanghai housing prices,” study the treatment effect of a major policy change on Shanghai’s housing market by employing panel data from March 2009 to December 2021. They use the panel data approach suggested by Hsiao et al. (2012) to estimate the treatment effects and a time series approach to disentangle the treatment effects and the effects of the pandemic. For time periods after the outbreak of the pandemic, they find no significant impact of the pandemic on the real estate price indices between 2020 and 2021.

Papke and Wooldridge, in their paper “A simple, robust test for choosing the level of fixed effects in linear panel data models,” propose a test that allows one to determine whether controlling for fixed effects at the more aggregate level is sufficient. The alternative is that one should allow for fixed effects at the unit level. The regression-based test is simple to carry out, even for unbalanced panels. In addition, the test is easily made robust to arbitrary heteroskedasticity, serial correlation across time, and even cluster correlation at the group level.

## 2 Stochastic frontiers analysis and efficiency/productivity measurement

Although not a formal contributor to the Special Issue, C. A. Knox Lovell was at the core of the original contributions by Peter on the topic of stochastic frontier analysis (SFA) and efficiency/productivity measurement. He has provided us with a bit of perspective in private correspondences on the development of SFA, for which of course he was also responsible. Knox has pointed out that Peter was more than an important contributor to the field; he was a creator of SFA, with Aigner et al. (1977) (Meeusen and van den Broeck, 1977, published a similar and nearly simultaneous paper in the *International Economic Review*, unknown at the time to Aigner et al. and similarly for Meeusen and van den Broeck) and directs us to the *Journal of Econometrics* 2023 50th Anniversary Jubilee Issue (2023) where the paper is reproduced along with the authors’ recollections of the origins and subsequent developments of SFA. Knox noted as well Peter Schmidt’s role in subsequent developments of SFA, in particular, the conversion of the production frontier to the cost frontier (Schmidt and Lovell 1979, 1980), methods to estimate the efficiency of individual producers (Jondrow et al. 1982), and extensions to panel models wherein efficiency is allowed to vary both across producers and through time (Schmidt and Sickles 1984; Cornwell et al. 1990). Among the many additional contributions Peter has made in this field of study, Lovell points out two additional seminal works that focus on “measurable and policy-sensitive determinants of efficiency,” Wang and Schmidt (2002) and Alvarez et al. (2006).

We now turn to brief discussions of each formal contribution to the special issue that relate to this aspect of Peter Schmidt’s work.

Chetty and Heckman, in “Internal adjustment costs of firm-specific factors and the neoclassical theory of the firm,” consider the predictions for factor demand of a two-sector vertically integrated model of firms producing output using firm-specific capital along with a second sector producing firm-specific capital that adapts raw capital purchased in the market. They find that aggregating over both sectors produces short-run and long-run factor demand functions that appear to be perverse, but when disaggregated they obey standard neoclassical properties. They conclude that adjustment costs create the appearance of static inefficiency in the presence of dynamic efficiency.

In “Proportional incremental cost probability functions and their frontiers,” Féve, Florens, and Simar suggest an alternative semi-parametric model that avoids the drawbacks of the two-stage methods of estimating frontier models. Their approach is based on a class of model called the proportional incremental cost functions, which is adapted from the Cox proportional hazard model. Use of this approach avoids the first-stage nonparametric estimation of the frontier and avoids the curse of dimensionality keeping the parametric rates of convergence for the parameters of interest.

Koenker, in “Hotelling tubes, confidence bands and conformal inference,” proposes using Hotelling’s tube methods for constructing nonparametric quantile regression confidence bands. Hotelling’s methods strengthen the performance of such bands. Koenker’s innovation is based on recent developments in conformal inference—considered to be a new approach to nonparametric inference for stochastic frontier models.

“In Indirect inference estimation of stochastic production frontier models with skew-normal noise,” Lai and Kumbhakar consider a stochastic frontier model in which both the noise and inefficiency components are asymmetric, viz., the noise term is skew normal and the inefficiency term is half normal. This formulation avoids the criticism that skewness of the composite error term (sum of the noise and inefficiency) cannot be an indicator of inefficiency because skewness can also arise from the noise term. They further generalize the model by introducing determinants of skewness of the noise term as well as determinants of inefficiency, and provide both simulation and empirical results using the indirect inference estimation approach.

In “The noise error component in stochastic frontier analysis,” Papadopoulos examines the relation between predicted noise and predicted inefficiency. For the Normal-Half Normal and the Normal-Exponential error specification, he provides its conditional expectation as a predictor and examines its distribution in relation to the marginal law. He also derives the conditional distribution of the noise and computes confidence intervals and the probability of over-predicting it.

In “An alternative corrected ordinary least squares estimator for the stochastic frontier model,” Parmeter and Zhao consider an extension of the corrected ordinary least squares (COLS) estimator for the stochastic frontier model. They propose a novel modification to COLS by using the first moment of the absolute value of the composite error term in place of the third moment for both the Normal-Half Normal and Normal-Exponential specifications. They demonstrate via simulations that this modification considerably reduces the occurrence of both Type I and Type II failures.

### 3 Time series

Ahn's "Likelihood based inference for dynamic panel data models" examines the asymptotic and finite sample distributions of maximum likelihood (ML) estimators for both stationary and nonstationary time series processes. He links the identification criteria developed by Sagan (1983) to the singularity of the information matrix for ML estimators under nonstationary. A major finding is that when data follow unit root processes without or with drift, the ML estimators are consistent, but they have nonstandard asymptotic distributions and their convergence rates are slower than  $n^{1/2}$ . Additionally, Monte Carlo experiments show that the modified LR tests are much better sized than the corresponding Wald tests. Although LR tests tend to slightly over-reject the unit root hypothesis in small samples they maintain good finite sample power properties.

In "Approximating long memory processes with low order autoregressions with implications for modeling realized volatility," Baillie, Cho, and Rho show that for realistic ranges of the long memory parameter in fractionally integrated times series, the ordinary least squares estimators of an AR(p) model will have nonstandard rates of convergence to nonstandard distributions. The AR parameter and impulse response function estimators will be of questionable value to researchers as well as models that use these estimators to represent realized volatility (RV) in financial markets.

In the topical empirical treatment "Does climate change affect economic data?," Choi derives seasonal factors based on US temperature, gasoline price, and fresh food price data sets using Kalman state smoothers and principal component analysis to show that seasonal volatilities have increased over the last four decades. Climate change is undoubtedly reflected in the temperature data. The three data sets show similar patterns from the 1990s, which suggests that climate change may have affected the price volatility behavior.

Phillips and Yu integrate flat trading features into an efficient price process in their modeling of asset price determination in their contribution "Information loss in volatility measurement with flat price trading." They develop a limit theory for the usual measure of integrated volatility, the conventional realized volatility (RV), and show that estimated RV, as well as estimated quarticity, has inflated asymptotic variances that depend on the probability of flat trading. Extensions to models with microstructure noise are also provided and the effect of flat trading using tick-by-tick data is empirically evaluated.

In "Forecasting in the presence of in and out of sample breaks," Xu and Perron use a frequentist-based paradigm, modified by modeling the probability of shifts, based on covariates that can also be forecasted and by a built-in mean reversion mechanism for the evolution of the parameters. Estimation is based on a mixture Kalman filter and a Monte Carlo expectation maximization algorithm and simulation results show that their approach is superior to standard forecasting models that are robust to model misspecifications. Empirical applications also are provided and illustrate the substantial gains in forecasting accuracy based on their new methods.

## 4 Applied econometrics, copulas, and nonparametric methods

The dependency structure of US commodity futures across sectors between 2004 and 2022 is the focus of “Multivariate models of commodity futures markets: a dynamic copula approach” by Chen, Li, Wang, and Zhang. During their study period, both the 2008 financial crisis and the COVID-19 pandemic caused substantial disruptions in markets. Their copula-based models address these major shocks by allowing for time-varying nonlinear and asymmetric dependence using flexible integrating elliptical and skewed copulas. Their empirical findings point to increasing connectedness among commodities during both of these events. Risk management strategies in commodity markets are also analyzed and they find that strategies using portfolio weights based on their dynamic copulas dominate those based on an equal-weighted portfolio.

Dang and Ullah use kernel regularized least squares (KRLS) in developing their two-step estimator of a nonparametric regression function in “Generalized kernel regularized least squares estimator with parametric error covariance.” The KRLS can model a very general covariance structure, similar to the HAC heteroskedastic and autocorrelation consistent covariance estimator, and the authors detail the construction of the estimator and also provide derivations of bias, variance, and asymptotics. Simulations and an empirical illustration provide both finite sample evidence of the appeal for the KRLS two-step estimator as well as its feasibility and relative ease of interpretability in terms of average partial effects of the inputs in their airline cost study, which illustrates the usefulness and feasibility of their new techniques.

Lahiri and Yang develop a new ensemble econometric model to predict binary outcomes in their paper, “Predicting binary outcomes based on the pair-copula construction.” Pair-copula construction (PCC) is used to optimally combine diverse information while allowing the conditional copula to depend on the conditioning variable non-parametrically. The authors use their new methods to predict US business cycle peaks, using Conference Board leading indicators. The predictive accuracy of their estimates using the receiver operating characteristic curve criteria is found to perform well in comparison with other widely used combination models. The appeal of their new ensemble predictor is highlighted by a number of diagnostic measures that point to different aspects of its advantages.

Varaku and Sickles utilize relatively new nonparametric methods and machine learning techniques in their paper “Public subsidies and innovation: a doubly robust machine learning approach leveraging deep neural networks.” They use Eurostat firm-level data to assess the effects of public subsidies on firms’ R&D input and output. Average treatment effects are estimated based on the selection on observables assumption as well as selection on unobservables that may also result in nonrandom subsidy assignment. Instrumental variables (IVs) are used to identify the local instrumental variable (LIV) curve. Identification of the LIV is obtained via double machine learning, combining IV estimation, neural networks, and deep neural networks to learn functional forms from the data. Major findings point to the positive and significant role that public subsidies have in increasing both R&D intensity and R&D output.

## 5 Limited dependent variables

Hirukawa, Liu, Murtazashvili, and Prokhorov focus on the Heckman (1979) sample selection model and utilize lasso-based criteria to select the control variables in their contribution, “DS-HECK: double-lasso estimation of Heckman selection model.” They point out that the usual lasso under-selects, adding the additional problem of omitted variable bias to the selection problem, and address the shortcoming of the lasso in this context by using a double lasso in the selection equation and in the calculation of the variance matrix. Simulations and a study of drivers of female US labor market participation and earnings are provided as well as a dedicated Stata procedure, **dsheckman**.

In “Simultaneity in binary outcome models with an application to employment for couples,” Honoré, Hu, Kyriazidou, and Weidner generalize the dynamic panel data model of Ahn and Schmidt (1995) to allow not only for lagged dependent variable and fixed effects in short panels, but also for bivariate outcomes studied by Schmidt and Strauss (1975) in their simultaneous logit model. Their estimators combine both conditional likelihood and method of moments methods and are used to examine the intra-household relationship in employment. Their major finding points to significant variations in within-household dependence in employment due to a couple’s ethnicity after controlling for unobserved household specific heterogeneity.

## 6 Closing Remarks

We owe a debt of gratitude to the contributing authors for trusting us to handle their research at the *Journal* rather than submit elsewhere, and patience for our many pleas to stay on top of revisions. On a final note, we would like to thank the many very capable referees who provided excellent criticism and feedback on the contributed papers. We thank them for their dedicated service in helping us put this special issue together and ensuring its high quality. We also owe a great debt of appreciation to the editorial board of *Empirical Economics* for approving the initial proposal for this issue.

## References

- Ahn SC, Schmidt P (1995) Efficient estimation of models for dynamic panel data. *J Econom* 68(1):5–27
- Ahn SC, Lee YH, Schmidt P (2013) Panel data models with multiple time-varying individual effects. *J Econom* 174(1):1–14
- Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6(1):21–37
- Alvarez A, Amsler C, Orea L, Schmidt P (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *J Prod Anal* 25(3):201–212
- Baltagi BH, Bresson G, Chaturvedi A, Lacroix G (2018) Robust linear static panel data models using -contamination. *J Econom* 202:108–123
- Baltagi BH, Bresson G, Chaturvedi A, Lacroix G (2021) Robust dynamic panel data models using -contamination. In: Chudik A, Hsiao C, Timmermann A (eds) *Advances in econometrics, essays in honor of M. Hashem Pesaran*. Emerald Publishing (forthcoming)

- Cornwell C, Schmidt P, Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. *J Econom* 46(1):185–200
- Drukker DM, Egger PH, Prucha IR (2022) Simultaneous equations models with higher-order spatial or social network interactions. *Econom Theory* (forthcoming)
- Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica* 47:153–161
- Hsiao C, Ching HS, Wan SK (2012) A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *J Appl Econom* 27(5):705–740
- Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econom* 19(2–3):233–238
- Journal of Econometrics* (2023) *Journal of econometrics jubilee issue—celebrating our fiftieth anniversary: 1973—2023*. Ng S, Tamer E (Eds.) 234(Supplement):1–120
- Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18(2):435–444
- Sagan JD (1983) Identification and lack of identification. *Econometrica* 51:1605–1634
- Schmidt P, Lovell CAK (1979) Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *J Econom* 9(3):343–366
- Schmidt P, Lovell CAK (1980) Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated. *J Econom* 3(1):83–100
- Schmidt P, Sickles RC (1984) Production frontiers and panel data. *J Bus Economic Stat* 2(4):367–374
- Schmidt P, Strauss RP (1975) Estimation of models with jointly dependent qualitative variables: a simultaneous logit approach. *Econometrica* 43(4):745–755
- Wang H-J, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *J Prod Anal* 18(2):129–144

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.