Editors' introduction: Identification, Simulation and Finite-sample Inference

M.-C. Beaulieu, J.-M. Dufour, L. Khalaf and A.C. MacKinlay

June 10, 2015

Introduction

Recent years have witnessed the enhancement of computational capacity available for research in finance and economics. One important development involves simulation-based procedures which have increased our ability to deal with identification problems and propose inference methods with better finite-sample properties. In this case, identification refers here to the possibility of distinguishing between different values of a parameter in a statistical or equilibrium-based model, given observable data. It is directly linked to the quality of the inference one can make about such parameters and its use for decision. In this special issue, contributors pay special attention to finite-sample inference, taking into account the fact that enlarged computational capacities help produce tools which can enhance the quality of inference. As such, these opportunities have led both to new research as well as new challenges. This special issue on *Identification, Simulation and Finite-sample Inference* aims at stimulating research and empirical analysis in this field.

Conceptual background

Identification is a wide concept. Here, we focus on the standard notion summarized in Dufour and Hsiao (2008). In economics and finance, one typically assumes that some (possibly theory-based) process or structure has generated or can represent observable data. However, statistical information on the parameters describing such structures can only be derived from the distribution of observations. In this context, identification refers to the possibility of: (i) determining parameters of interest from observable data, and (ii) drawing inferences on the underlying theoretical structure. The latter need not be structural in a strict sense, since reduced forms are also covered by this definition.

This application requires maintaining a reasonable balance between structural modeling based on (often highly stylized) theoretical equilibrium models, and statistical approaches that try to represent the actual dynamics of economic and financial data. Furthermore, model concordance with observations should be assessed, before predictions are employed for decision or policy. This fact stresses the importance of "robustness" considerations in both theoretical and empirical analysis, as well as policy decisions. Another aspect of "robustness" analysis consists in assessing what can and cannot be done (such as parameter estimation, forecasting, and policy

analysis) on the basis of models which may not be well "identified". In this context, it is indeed significant that some of the above ideas are highlighted in the work of the 2011 and 2013 recipients of the Nobel Prize in Economics: the importance of the concept of robustness has been recently emphasized by Hansen and Sargent [Hansen and Sargent (2007), Robustness, PUP, http://press.princeton.edu/titles/8535.html], while the possible unreliability of badly identified structural models was a major motivation of the work of Sims (1980, Econometrica).

On simulation, the reader will find different applications explored by the authors participating in this special issue. These cover broad problems such as density estimation, standard error corrections, shrinkage estimation, reduced rank regression, non-stationarity and structural breaks, as well as specific finance applications such as portfolio and market efficiency, predictability, and factor pricing. Simulation-based and resampling methodologies are changing economics and finance, offering economists and financial analysts the opportunity to examine many non-standard problems with greater generality and precision. Indeed, resampling methods (Monte Carlo tests, bootstrapping, susbsampling, jacknife) are computationally intensive, which makes them easier to apply in an environment with greater access to computing power. While simulation is powerful and relatively easy to apply, it is not without pitfalls. If researchers and practitioners wish to make full use of the power of simulation-based statistics, they must have a firm grasp of their key statistical foundations.

Ultimately, better identification using simulation methods will lead to better inference. This is interesting for policy makers, since the methods presented here improve the quality of inference. The aim is to set forth novel simulation-based and robust methodologies, and to assess their value as inference tools in economics and finance. Improving finite-sample accuracy is a major concern, so practitioners in economics and finance can use state-of-the-art simulation-based inference procedures appropriately and effectively.

Contributions

The range of applications presented in this special issue is wide: the classic mean-variance efficiency problem in finance (Gungor and Luger), market efficiency (Taamouti), the identification of important pricing factors and macroeconomic variables (Beaulieu, Dufour and Khalaf, Kichian, and Takongmo andStevanovic) as well as income distribution studies (Charpentier and Flachaire). Although most articles focus on policy-relevant econometric problems, papers with direct policy implications include Kichian on the natural employment rate, and Charpentier and Flachaire on inequality analysis.

A number of econometric contributions are included in this special issue. Beaulieu, Dufour and Khalaf as well as Kichian propose identification-robust inference methods, while contributions to bootstrapping include extensions and applications in various papers. MacKinnon presents wild cluster bootstrap confidence intervals, while Bergamelli-Novotny-Urga develop entropy-based block bootstraps for non-stationary processes. Gungor and Luger exploit the double bootstrap for multiple testing. Taamouti and Charpentier and Flachaire focus on nonparametric testing

and estimation. In the article by Taamouti, sign transformations are considered, while Charpentier and Flachaire use kernel techniques. Finally, this special issue also offers contributions to deal with dimensionality problems: mean-variance efficiency tests using assets rather than portfolios (Gungor and Luger), shrinkage instrumental variable estimators for large datasets (Carriero, Kapetanios and Marcellino), and structural breaks in factor analysis (Takongmo and Stevanovic).

MacKinnon (**Wild Cluster Bootstrap Confidence Intervals**) studies an important standard error correction which is widely adopted to build confidence intervals controlling for clustering, although its reliability even with large enough samples can be questionable. As an alternative to standard Wald-type based intervals, this article proposes to invert Wild Bootstrap based improved tests. Simple transformations as well as distributional approximations are also proposed. Dramatic finite-sample improvements are demonstrated, specifically when the number of clusters is small.

Gungor and Luger (**Bootstrap tests of mean-variance efficiency with multiple portfolio groupings**) study mean-variance efficiency tests using assets rather than portfolios. The proposed methods exploit multiple test approaches which are gaining popularity in empirical finance. The double bootstrap is used to correct for singularities arising from degrees-of-freedom shortages as the number of assets grows relative to time-series length. In view of the emerging literature that revisits portfolio based asset pricing tests this paper provides an innovative and promising approach.

Carriero, Kapetanios and Marcellino (**A Shrinkage Instrumental Variable Estimator for Large Datasets**) study instrumental variable estimation with many instruments. The paper exploits a formal shrinkage strategy to propose a new, intuitive and analytically tractable two-stage least-squares-type estimator. Consistency is proved under general conditions allowing for the number of instruments to be large relative to the total number of observations. Again, a comprehensive simulation experiment documents sizeable finite-sample improvements for realistic designs.

Taamouti (**Finite-Sample Sign-Based Inference in Linear and Nonlinear Regression Models with Applications in Finance**) provides a survey of recently proposed orthogonality sign-tests. These include robustness in large and small samples to heteroskedasticity of an unspecified form, non-normality, and inexistence of moments. Because of their nonparametric foundations, the tests do not require complete models. The paper also presents applications of these methods to multi-horizon stock return predictability and to market efficiency tests. Such a survey of theory as well as application is as yet unavailable, despite the strong advantages of such methods.

Bergamelli, Novotny and Urga (**Maximum Non-Extensive Entropy Block Bootstrap for Non-stationary Processes**) study the bootstrap with non- stationary data. The paper revisits recent entropy-based methods, documents their failure in this case, and propose blocking-schemes to correct such failures. The latter specifically focus on fat-tailed distributions in view of financial

applications. An extensive simulation study focuses on the well-known Dickey-Fuller test which despite its popularity still suffers from enduring size and power problems in finite samples.

Charpentier and Flachaire (**Log-transform kernel density estimation of income distribution**) focus on the nonparametric estimation of income distributions. Typically, such distributions are heavy-tailed, so kernel estimation does not work well even with the very large samples often available in income distribution data. The paper proposes a very simple transformation, which upon using simulation methods, allows one to sharply improve finite-sample fit without distorting the specificities of income analysis such as the assessment of poverty or inequality.

Kichian (**Identification-Robust Estimates of the Canadian Natural Rate of Interest**) focuses on the estimation of one of the most important monetary-policy measures, the so-called natural rate of interest. It is well known that despite widespread central banking relevance, endogeneity, under-identification, and errors-in-variables problems have rarely been addressed in this context. This paper controls for estimation uncertainty despite all the above concerns. Various contemporaneous and forward-looking interest rate rules are considered as a foundation for empirical analysis, as well as central bank real-time staff projections, which enhances the policy relevance of the method. The paper focuses on the Canadian case on which available work is scarce, despite the leading role of the Canadian central bank during the recent financial crisis.

Takongmo and Stevanovic (**Selection of the number of factors in presence of structural instability: a Monte Carlo study**) analyse factor models in the presence of structural instability. Concrete guidance on selection methods is proposed in view of data-rich macroeconomic and financial studies. Modeling structural instability via time-varying factor loadings as proposed is a novel, tractable and promising technique that can provide an explanation for some long lasting empirical puzzles in macro-finance.

Finally, Beaulieu, Dufour and Khalaf (**Identification-Robust Factor Pricing: Canadian Evidence)** analyze factor models based on the Arbitrage Pricing Theory (APT) by analytically inverting Hotelling-type pivotal finite-sample tests. This approach may be viewed as a multivariate and multi-factor extension of the Fieller method. Results are applied to multi-factor asset pricing models with Canadian data, on which empirical related work is scarce. Despite evidence of weak identification, results support a three-factor model (Fama-French-type) after 2000. The momentum factor severely compromises identification, which calls for caution in interpreting existing work on momentum effects in the Canadian market.

Challenges

The papers in this special issue propose new solutions to enduring research questions. A common denominator across all contributions is a methodology well-grounded in statistical

theory, leading to more reliable procedures in models for which only standard large-sample methods are typically available. In addition to convincing documented successes, communalities across papers suggest current challenges. Among these, the following ones are in our view especially relevant: (i) data-rich inference problems calling for information reduction and/or organisation of data [e.g. many instruments, many factors, portfolio formation], (ii) methods that are valid in finite samples under weak distributional assumptions allowing in particular for heavy tails, instabilities and heterogeneity, and (iii) attention to identification particularly although not exclusively with equilibrium-based theory founded models. Such challenges outline interesting avenues for future research.