Statistical Challenges in Modern Economics

Wei Luo[1,2,3] Yonghong Zhou[1]
[1] Jinan University, Guangzhou, China
[2] Guangdong University of Technology, Guangzhou, China
[3] Corresponding author: Wei Luo, e-mail: luowei80@gdut.edu.cn

Abstracts

Modern economics have the nature of quantity and have adopted empirical studies widely. Statistics is keen to devise ways to provide research access to modern economics. This paper seeks to review and clarify the role of statistics in the methodology perspective of economic analysis. Although the methods of econometrics are now well established in statistical literature, the application of these methods has posed us with considerable challenges considering the peculiarities of economic data and the time-variant of economic system. This presentation will show how these economic properties invalidate many of the common statistical approaches used in economic analysis and introduce the recent work to overcome them. It also highlights statistical adequacy as a primary criteria in econometrical models evaluation.

Keywords: econometrics, economic change, economic data, statistical adequacy

1. Introduction

Statistics have had deep impact on modern economics. At the same time, the credibility of statistics has been subjected to debate for years, which is partly due to the statistical tools themselves and mainly because of the uncertain nature of economic relations. Modern economics have the nature of quantity and have adopted empirical studies widely. The scientific and credible principle along with many new economic problems has created golden opportunities and significant challenges for the development of statistics. The first breakthrough was the rising of descriptive statistics in the late 19th century as more statistical observations occurred on economic variables available and urgent demands of quantitative analysis rose in economics. Landmarks included Engel curve and Edgeworth equation. Statistical tools such as correlation and regression analysis were taken up in economics by Galton and Pearson. Successive effort of Moore (1914, 1917), Wright (1927), Working (1928), Tinbergen (1929-30), Frisch (1933), and Schultz (1938), laid the basis of "statistical economics", the precursor of econometrics. In 1930s, the applicability of the probability calculus, the newly developed sampling theory and the significance tests on the analysis of economic data over econometric method led to high skeptic, which finally boosted the second revolution of statistics application to empirical research. It was Haavelmo (1944) who established a unified econometrics framework based on probability theory and statistics. Since then, model-based economic empirical analysis has become the

main stream research method. Statistical estimation methods such as likelihood estimation and moment estimation were largely used for identification, estimation and inference in econometrics. Simultaneously, Yule (1926), Wald (1938), Durbin and Watson (1950, 1951), Fisher (1955), Neyman and Pearson (1956), Savin (1973) developed appropriate test statistics of hypotheses. A notable development of Bayesian inference in econometrics and statistics made a significant step towards dynamic specification since 1970s.

The past three decades have witnessed the remarkable changes of world economic environment. No economic theories were ever testable, and many previous economic models or analysis tools were proved to invalid. For example, Sims (1980) queried the adequacy of simultaneous equations for their invoked identification assumptions. Hendry (1980) was highly skeptical of the effectiveness and validity of applying econometric methods. Nelson and Plossor（1982）found most macroeconomic time series were unit root. Spanos (1986) first separated misspecification testing from hypothesis testing. The development of microeconometrics, financial econometrics and the availability of high quantity economic data have posed new problems and challenges in statistics focusing on statistical adequacy.

Our focus is on the economic modeling founded with available data, using careful statistical analysis to guide the general model specification. We also highlight the potential problems that are related to economic data and economic structure changes and the recent work to overcome them. Considering the length restriction, this review has to leave out a wide range of important work inevitably.

In section 2, we introduce the assumptions and fundamental premises of modern economics in detail, and the statistical implication of stochastic disturbance, presenting both theory and examples. Section 3 discusses the impact of economic data peculiarities and economic change to statistics in economic analysis; some recent work to overcome them will be provided. A summary follows in section 4.

2. Statistical exploration in modern economics

Hendry (1995) exploited the assumptions and fundamental premises of modern economics on two grounds: first, economic system is treated as a stochastic process from a complete probability space. Second, economic data is the realization of this data generating process (DGP). We need to metamorphose the probability space into a less abstract probability models for realistic purposes. In practice, we get a finite sample of economic system $X=\left(X_1, X_2, \ldots, X_n\right)$, such as the growth rate of the economy. The parametric probability model $\Phi=\left\{f\left(x;\theta\right), \theta\in\Theta, x\in\Re_x\right\}$ specifies a family of densities $f\left(x;\theta\right)$ defined over the range of values ($\Re_x$) of the random variable X; a density function that constitutes the basic element of the stochastic process for each value of the parameter θ, as the latter varies over its range of values Θ: the parameter space. A series of partitioning and reduction are put on stochastic vector to gain an operational probability model. For example, joint

distribution model is reduced to conditional distribution model, and then to a conditional expectation model, that is population regression model. Under appropriate economic hypothesis and stochastic disturbance specification, the latter dominate model specification, parameter estimation and hypothesis testing in econometric analysis; probability models can identify economic quantitative relationship from the cumulative distribution. Considering the incomplete economic data may mismatch with hypothesis, it was realized that the knowledge created by any rule of generalizing from the particular is uncertain. Rao (1997) summarized the statistical solution to gain usable knowledge in the way which uncertainty can be quantified and expressed. Accordingly, economic models including stochastic disturbance specification can measure confidence level effectively in economical specific test. It is recognized that model specification is the basis of statistical inference. Pesaran and Smith (1985), Geweke et al., (2006), Spanos (2010) proposed statistical adequacy, when the economical and statistical specification of model conforms the regularities in data, as primary criteria in econometrical model evaluation.

3. Challenges of economic data and economic system

We now outline certain problems of various frontiers of research to illustrate the challenges of economic data and economic structural change. Some solutions to these problems will be provided.

3.1 Economic data issues

Data collection and empirical stylized facts summary are the starting point of modern economics. On one hand, technological innovations help us to collect massive amount of high quantity data with relatively low cost. On the other hand, the peculiarities of massive economical data along with new scientific problems have reshaped statistical thinking, data analysis and theoretical studies.

Firstly, modern statistics has basically focused on procedures and methods appropriate for the analysis of experimental and sample survey data. However, economic data are not the results of repeated randomized trial but rather obtained as by-products of business and government activities. Therefore, it is impossible to filter out the effect of the factors other than economic models from observed economic data, which will cause difficulties in economic empirical research. Design-based experimental or quasi-experimental research may help study economic man's behavior under controlling conditions, but experiments are expensive, and may not always be practical. Such a large-scale experiment to economic entity like China is beyond imagination. Since it is difficult to gain economic data from a well-designed trial, we then turn our attention to model-assisted or model-based approach to choose good designs and efficient design-unbiased or design-consistent estimators. So specification testing would be crucial to non-experimental type data. Neglecting specification test or expanding test statistics without considering the probabilistic assumptions of economic data may threaten the adequacy of statistical inference. For example, to test if a Cobb-Douglas production function $F\left(L,K\right)=AL^{a}K^{b}$ has the constant return to scale (CRS) nature, the null hypothesis is $H_0:a+b=1$. T-test may be invalid when

cross-section observes are conditional heteroscedasticity. New tests such as White test are needed. Residual-based tests for cointegration regression in longitudinal data presented by Kao (1999) cannot be derived directly from a Dickey–Fuller (DF) tests or an augmented Dickey–Fuller (ADF) test in time-series because the null distribution of residual-based cointegration tests depends on the asymptotics of least-squares dummy variable estimator. Monte Carlo experiments showed the latter would reject the null hypothesis excessively. Spanos (2010) doubt the validity of the Akaike-type Information Criterion (AIC) selecting parsimonious statistical models for assuming away the problem of model validation and ignoring the relevant error probabilities. Specification test is a field worth exploring: from Hausman test(Hausman,1978), heteroscedasticity test (Pagan and Hall, 1983)，autocorrelation test (Arellano and Bond, 1991), Wald test (Cragg and Donald, 1993), balancing test (Dehejia and Wahba, 2002), reduced rank test (Kleibergen and Paap, 2006) to nonlinear and nonstationary models (Wang and Phillips, 2012), we expect nonstationary and nonlinearity in statistical inference can be used as testing instruments; nonlinear transformations of integrated time series have some important statistical properties that can be exploited to develop tests with some desirable characteristics.

Next, the statistical methods available are generally suitable for large samples while the economic data typically have a rather limited coverage. For example, the GDP growth rate of China in last decade:

Y_{2002}	Y_{2003}	Y_{2004}	Y_{2005}	Y_{2006}	Y_{2007}	Y_{2008}	Y_{2009}	Y_{2010}	Y_{2011}
9.1%	10.0%	10.1%	9.9%	10.7%	14.2%	9.6%	9.1%	10.4%	9.2%

GDP growth rate in each year is regarded as a different stochastic variable, and each stochastic variable has only one observation. Usually, the observed data are handled in such a manner to permit as much time- stationary or homogeneity as possible, so GDP growth rate in each year is regarded from identical distribution. However, it's difficult to make specification test of the prerequisite in practice. Small area estimation generally used for small sample sizes through auxiliary information such as administrative records and census may provide a new orientation.

Thirdly, modern microeconomics benefits a lot from the expansion of electronic recording and collection of data at individual level. Microeconomic data are typically collected using cross-section and longitudinal surveys. Microeconomic data are usually enormous and of different types, which are subjected to problems of complex survey methodology and simple selection bias. Microeconomic analysis based on discrete or censored individual-level data, in which case nonlinear methods such as logit, probit, and Tobit models are used. Dealing with such issues leads to statistical inference based on more difficult asymptotic theory and robust statistical inference. Recently, important advances have been made in longitudinal data, including: analysis of longitudinal data using Linear mixed models (Verbeke and Molenberghs, 2009), latent variable models (Muthen, 2011), cross-lagged structural equation models (Vicente and Salgueiro, 2013) and shape regression(Fishbaugh and Durrleman, 2012); adjusting for measurement error in the value added model (Ferrão and Goldstein, 2009); specifying the probability distribution inference under informative sampling

(Eideh and Nathan, 2009).

Finally, the pool quality of sensitive data and declining response rate may be solved by a substantial reliance on administrative data; and a more sophisticated interpretation of survey data when administrative records are unavailable.

Researchers also focus on the problems of clustering of observations, specification tests and model selection, exact and near-exact distributions for the test statistic, sensitivity and specificity in compound tests, estimation of the extreme value index, semi-parametric and nonparametric models, measurement errors, missing data, and aggregation over time.

3.2 time-variant issues

The correlations of economic system are usually time-variant, accompanied by structural change. The economic system with time-variant is a nonstationary process, that means the rule of economic operation has different characteristics or stylized facts with the change of time, and simple aggregation of data cannot be used for statistical analysis. The time-variant of economic system will jeopardize out-of-sample forecasts and effective economic policy decision making.

There are two essential types of time-variant in economic system, one is occasional structural breaks and the other is smooth structural breaks. Occasional structural break when the break date is known with enough sample data after break, were well discussed in Sibbertsen (2004). The classical Chow test is widely used for single structure change, but it may be uninformative or misleading as the unknown break date is endogenous and the test is likely to indicate a break falsely when none in fact exists.

The Change of Smooth structural assumes economic system has evolutionary time-variant. Because the economic behavior may need a period to accommodate abrupt break, or aggregation of individual occasional breaks may perform macro evolutionary. Smooth structural change is common in economic, which creates great difficulties in applied research since it's not easy to distinguish either the long memory property from the smooth break process or structural time-variant from the structural nonlinear.

Most macroeconomic data are aggregated by microeconomic data; however, the aggregation may change economic correlations. For example, considering variables I_{it} and C_{it} are income and consumption of the i-th family during the t-th year, the family consume function can be written as

$$C_{it} = \alpha_i + \beta_i I_{it} + \varepsilon_{it}, \quad i = 1,\ldots,n, \quad t = 1,\ldots,T .$$

where β_i represents marginal propensity to consume of i-th family. The expressions $I_t = \sum_{i=1}^{n} I_{it}$ and $C_t = \sum_{i=1}^{n} C_{it}$ represent the general income and consumption in the t-th year. We than have

$$C_i = \alpha + \beta_l I_t + \varepsilon_t, \quad t = 1,\ldots,T$$

Where $\beta_l = \sum_{i=1}^{n} \beta_i I_{it} \Big/ \sum_{i=1}^{n} I_{it}$ represents general marginal propensity of consume.

Unless all families have homogeneous income growth, parameter β_l is time-variant. The aggregation lead macro consumes function nonlinear.

Statistical literature contains quite a lot of work in structural change, including the following innovations: tests for structural changes under very general conditions (Sakoulis et al., 2010); methods to select the number of breaks (Verbesselt et al., 2010); consistency of estimates structural changes under finite sample (Perron and Yamamoto, 2011); confidence intervals in various hypotheses about the structure of the data (Ponciano et al., 2009); Bayesian, bootstrap, nonlinear and nonparametric methods to estimating and detecting changes.

Currently, under progress is trying to address these issues: limit distributions of estimates of break dates in a cointegrated system with multiple structural changes, issues of nonmonotonic power functions for tests of structural change, evaluating the frequency of permanent shocks issues, distinguishing between long-memory processes and short-memory processes with structural changes, and their inner-relationship.

4. Conclusions

Nowadays, statistics is experiencing a credibility revolution with adequacy of the postulated statistical model, which captures the statistical systematic information contained in the data. Thus, statistical adequacy is the primary criteria in econometrical models evaluation; while better data and more robust estimation methods are the other part of the story; the emphasis gradually shifted from estimation and inference based on a given tightly parameterized specification to diagnostic testing, specification searches. Just as Hendry (1980) declared: "The three golden rules of econometrics are test, test and test", specification test is crucial to economic modeling. We hope that this paper will provide a valuable benchmark to direct research in specific directions and to alert researchers of the potential merits and drawbacks of various approaches.

References

Angrist, J. D. and J. S. Pischke (2010) "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3-30.

C. R. Rao (1997) "Statistics and truth putting chance to work," World Scientific Publishing Co. Pte. Ltd, Singapore.

Haavelmo, T. (1944) "The probability approach in econometrics," *Econometrica*, 12, Supplement, 1-118.

Hendry, D. F. (1980) "Econometrics—Alchemy or Science?" *Economica*, 47, 387-406.

Pesaran, H., and Smith, R. (1985) "Evaluation of Macroeconometric Models," *Economic Modeling*, 2, 125-134.

Sibbertsen, P. (2004) "Long memory versus structural breaks: an overview," *Statistical Papers*, 45, 465-515.

Spanos, A. (2010) "Akaike-type Criteria and the Reliability of Inference: Model Selection versus Statistical Model Specification," *Journal of Econometrics*, 158, 204-220.