"Refitted Cross-Validation in Ultra High Dimensional Regression"
Jianqing Fan (Princeton University, USA), Shaojun Guo (Institute of Applied Mathematics, Academia Sinica, Taiwan), Ning Hao (Princeton University, USA)
Abstract:
Variance estimation is a fundamental problem in statistical modeling. In ultrahigh dimensional linear regressions where the dimensionality is much larger than sample size, traditional variance estimation techniques are not applicable. Recent advances on variable selection in ultrahigh dimensional linear regressions make this problem more accessible. One of the major problems in ultrahigh dimensional regression is the high spurious correlation between the unobserved realized noise and some of the predictors. As a result, the realized noises are actually predicted when extra irrelevant variables are selected, leading to serious underestimate of the noise level. In this paper, we propose a two-stage refitted procedure via a data splitting technique, called refitted cross-validation (RCV), to attenuate the influence of irrelevant variables with high spurious correlations. Our asymptotic results show that the resulting procedure performs as well as the oracle estimator, which knows in advance the mean regression function. The simulation studies lend further support to our theoretical claims. The naive two-stage estimator which fits the selected variables in the first stage and the plug-in one stage estimators using LASSO and SCAD are also studied and compared. Their performances can be improved by the proposed RCV method. The methods are applied to assess the forecasting errors of home price indices in the core based statistical areas in the US.

"Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain"
Alexandre Belloni, Daniel Chen, Victor Chernozhukov (MIT, USA), Christian Hansen
Abstract
We develop results for the use of LASSO and Post-LASSO methods to form first-stage predictions and estimate optimal instruments in linear instrumental variables (IV) models with many instruments, p, that apply even when p is much larger than the sample size, n. We rigorously develop asymptotic distribution and inference theory for the resulting IV estimators and provide conditions under which these estimators are asymptotically oracle-efficient. In simulation experiments, the LASSO-based IV estimator with a data-driven penalty performs well compared to recently advocated many-instrument-robust procedures. In an empirical example dealing with the effect of judicial eminent domain decisions on economic outcomes, the LASSO-based IV estimator substantially reduces estimated standard errors allowing one to draw much more precise conclusions about the economic effects of these decisions. Optimal instruments are conditional expectations; and in developing the IV results, we also establish a series of new results for LASSO and Post-LASSO estimators of non-parametric conditional expectation functions which are of independent theoretical and practical interest. Specifically, we develop the asymptotic theory for these estimators that allows for non-Gaussian, heteroscedastic disturbances, which is important for econometric applications. By innovatively using moderate deviation theory for self-normalized sums, we provide convergence rates for these estimators that are as sharp as in the homoscedastic Gaussian case under the weak condition that $\log p = o(n^{1/3})$. Moreover, as a practical innovation, we provide a fully data-driven method for choosing the user-specified penalty that must be provided in obtaining LASSO and Post-LASSO estimates and establish its asymptotic validity under non-Gaussian, heteroscedastic disturbances.

"Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model"
Søren Johansen (University of Copenhagen and CREATES, Aarhus University), Morten Ørregaard Nielsen (Queen's University, Canada, and CREATES, Aarhus University, Denmark)
Abstract:
We consider model based inference in a fractionally cointegrated (or cofractional) vector autoregressive model based on the conditional Gaussian likelihood. The model allows the process X{t} to be fractional of order d and cofractional of order d-b; that is, there exist vectors β for which β′X{t} is fractional of order d-b. The parameters d and b satisfy either d≥b≥1/2, d=b≥1/2, or d=d0≥b≥1/2. Our main technical contribution is the proof of consistency of the maximum likelihood estimators on the set 1/2≤b≤d≤d1 for any d1≥d0. To this end, we consider the conditional likelihood as a stochastic process in the parameters, and prove that it converges in distribution when errors are i.i.d. with suitable moment conditions

and initial values are bounded. We then prove that the estimator of β is asymptotically mixed Gaussian and estimators of the remaining parameters are asymptotically Gaussian. We also find the asymptotic distribution of the likelihood ratio test for cointegration rank, which is a functional of fractional Brownian motion of type II.

"Classical Time-Varying FAVAR models - Estimation, Forecasting and Structural Analysis"

Sandra Eickmeier (European Cenral Bank and Deutsche Bundesbank, Germany), Wolfgang Lemke (Deutsche Bundesbank, Germany), and Massimiliano Marcellino (European University Institute, Italy and Bocconi University, Italy)

Abstract.

We propose a classical, Kalman filter-based, approach to analyze factor-augmented vector autoregressive (FAVAR) models with time variation in the factor loadings, in the factor dynamics, and in their conditional variance-covariance matrix. When the time-varying FAVAR is estimated using a large dataset of US variables, the results indicate minor changes in the factor dynamics and more marked variation in their volatility and their impact on key macroeconomic variables. Forecasts from the time varying FAVAR are substantially more accurate than those from a constant parameter FAVAR, especially for several inflation, financial and monetary variables, also over the more recent period. Finally, when the time-varying FAVAR is used to assess how monetary transmission to the economy has changed, we find substantial time-variation in the volatility of monetary policy shocks and their transmission to real activity and inflation.

"Aggregation in Large Dynamic Panels"

Hashem M. Pesaran (Cambridge University and CIMF, UK, and USC, USA), Alexander Chudik (European Central Bank, Germany and CIMF, UK)

Abstract:

This paper reviews Granger's contribution to the literature on aggregation and extends it by considering the case of large linear dynamic panels, where various interconnections between individual units are allowed. We notably permit distributed lags of all micro units to enter the individual micro relations, relax the independence assumption of the micro-distributed lag coefficients, and allow a general pattern of cross section dependence of micro innovations, which can be either strong or weak. Using the Pesaran's (2003) forecasting approach to derive the optimal aggregate model, our paper derives the optimal aggregate model of a factor-augmented VAR in N cross section units. The paper also discusses the aggregation error in this set-up, identify some of the distributional features of micro-parameters from aggregate relations, and proves Granger's (1980) conjecture about long memory properties of aggregate variables from a large scale dynamic econometric model. Monte Carlo experiments assess how these aggregate functions perform in small samples.

"Selecting the Correct Number of Factors in Approximate Factor Models: The Large Panel Case with Bridge Estimators"

Mehmet Caner (North Carolina State University, USA)

Abstract.

This paper proposes Bridge estimators to correctly select the number of factors in an approximate factor model when there are latent factors. This contributes to both econometrics and statistics literature. Instead of the information based criterion penalty in econometrics, we propose a penalty based on factor loadings. So this extends the Bridge estimators in least squares context to factor models. This is a new approach in factor models, we show that oracle property of Bridge estimators is preserved, and hence we can select the correct number of factors through high dimension reduction. Simulations show that our technique can do a better job than information based criterion on both autocorrelated and cross section correlated data. A simple example about US macro factors in the last 25 years is supplied in the paper as well.

"Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data"

Marta Banbura (European Central Bank, Germany) and Michele Modugno (European Central Bank, Germany)

Abstract

In this paper we propose a methodology to estimate a dynamic factor model on data sets with an arbitrary pattern of missing data. We modify the Expectation Maximisation (EM) algorithm as proposed for a dynamic factor model by Watson and Engle (1983) to the case with general pattern of missing data. We also extend the model to the case with serially correlated idiosyncratic component. The framework allows to handle efficiently and in an automatic manner sets of indicators characterized by different publication delays, frequencies and sample lengths. This can be relevant e.g. for young economies for which many indicators are compiled only since recently. We also show how to extract a model based news from a statistical data release within our framework and we derive the relationship between the news and the resulting forecast revision. This can be used for interpretation in e.g. nowcasting applications as it allows to determine the sign and size of a news as well as its contribution to the revision, in particular in case of simultaneous data releases. We evaluate the methodology in a Monte Carlo experiment and we apply it to nowcasting and backdating of euro area GDP.

"Factor Based Identification-Robust Inference in IV"

George Kapetanios (Queen Mary, University of London, UK), Lynda Khalaf (Carleton University, Canada), M. Marcellino (EUI and Bocconi University, Italy)

Abstract:

Weak-instruments robust methods raise important size/power trade-offs resulting from omitted instruments biases. The popular Anderson-Rubin test has the right size when the underlying first-stage model (that is, the model linking the structural equations right-hand-side endogenous variables to available instruments) is closed or is incomplete. Alternative methods are available that may outperform this statistic assuming a closed first-stage specification (that is, assuming that all instruments are accounted for). We show that information-reduction methods provide a useful and practical solution to this problem. Formally, we propose factor-based modifications to three popular weak-instruments-robust statistics, and illustrate their validity asymptotically and in .nite samples. Results are derived using asymptotic settings that are commonly used in both the factor-analysis and weak-instruments literatures. For the Anderson-Rubin statistic, we also provide analytical finite sample results under usual assumptions. An illustrative Monte Carlo study reveals the following. (1) Our factor-based corrections circumvent the size problems resulting from instrument omissions and improve the power of the Anderson-Rubin statistic. (2) Once corrected through factor reduction, all considered statistics perform equally well. Results suggest that factor-reduction holds promise as a unifying solution to the many instruments problem.

"Testing Rates: Nonparametric Nonstationarity Tests"

Federico Bandi (The Johns Hopkins Carey Business School, Baltimore, USA) and Valentina Corradi (Warwick University, UK)

Abstract:

We propose nonstationarity tests which exploit the different divergence rates of the occupation measures of a (possibly nonlinear) process under the null of nonstationarity (stationarity) versus the alternative of stationarity (nonstationarity). We consider both discrete- and continuous-time series. The discrete-time case covers Harris recurrent Markov chains (as in, e.g., Karlsen and Tjostheim, 2001, Guerre, 2004, and Schienle, 2008) and (near-)integrated processes (as in, e.g., Wang and Phillips, 2009a, 2009b). The continuous-time case focuses on diffusion processes. In addition to their simplicity and reliance on tabutaled critical values, we show that the size and power properties of the proposed tests are satisfactory and generally superior to those of traditional approaches to nonstationarity testing.

"Model Selection in Equations with Many `Small' Effects"

Jennifer L. Castle (Oxford University, UK) and David F. Hendry (Oxford University, UK)

Abstract:

Model selection from a general unrestricted model (GUM) can potentially confront many small relevant effects that may not be retained by selecting individual variables, but can be partially captured by latent factors, together with important individual determinants, while eliminating many irrelevant variables. We exploit the ability of automatic model selection procedures to handle perfect collinearity and more candidate variables, N, than observations, T to investigate including both the candidate regressors and all their principal components. This allows our approach to embed many salient explanatory variables, their lags, functional

form transformations, and multiple location shifts together with `factor' structures, without being concerned that the resulting N>T. The factors might enter as `genuine' explanatory variables, or as approximations to the many small relevant effects that would not be retained individually. Potential applications to forecasting are noted.