Workshop on panel data

University of Groningen & University of Amsterdam

Friday afternoon and Saturday morning 5-6 April 2019

Venue: De Burcht, Henri Polaklaan 9, Amsterdam
Format: 45 minutes per speaker, 2 or 3 presentations in a session
Registration: free of charge, by email (events-ase@uva.nl) before 1 April 2019
Website: www.ase.uva.nl/panel-data
Organizers: Maurice Bun (DNB/UvA) & Arturas Juodis (RUG)

Programme Friday 5 April

13:00-13:45 Welcome

13:45-16:00 Session 1 (Chair: Peter Boswijk)

Eleni Aristodemou (University of Amsterdam)
Semiparametric identification in panel data discrete response models
Discussant: Geert Dhaene (KU Leuven)

Luis Candelaria (University of Warwick)
A semiparametric network formation model with multiple linear fixed effects
Discussant: Arturas Juodis (University of Groningen)

Chris Muris (University of Bristol)
Binarization for panel models with fixed effects
Discussant: Martin Carree (Maastricht University)

16:00-16:30 coffee/tea break

16:30-18:00 Session 2 (Chair: Frank Kleibergen)

Andreas Dzemski (University of Gothenburg)
Confidence set for group membership
Discussant: Otilia Boldea (Tilburg University)

Koen Jochmans (University of Cambridge)
Testing for correlation in error-component models
Discussant: Helmut Farbmacher (Munich Center for the Economics of Aging)

18:30-22:00 drinks/dinner (The Birdhouse, Plantage Middenlaan 46)

Programme Saturday 6 April

9:30-11:00 Session 3 (Chair: Kees Jan van Garderen)

Geert Mesters (Universitat Pompeu Fabra)
Nonlinear dynamic factor models with interacting level and volatility
Discussant: Noud van Giersbergen (University of Amsterdam)

Frank Windmeijer (University of Bristol)
Weak instruments, first-stage heteroskedasticity and the robust F-test
Discussant: Tom Wansbeek (University of Groningen)

11:00-11:30 coffee/tea break

11:30-13:45 Session 4 (Chair: Maurice Bun)

Joerg Breitung (University of Cologne)
Estimation of heterogeneous panels with systematic slope variations
Discussant: Julia Schaumburg (VU Amsterdam)

Lorenzo Trapani (University of Nottingham)
Determining the dimension of factor structures in non-stationary large datasets
Discussant: Andreas Pick (Erasmus University Rotterdam)

Gerdie Everaert (Ghent University)
Testing the rank condition for CCE-type estimators
Discussant: Hande Karabiyik (VU Amsterdam)

N.B. lunch (sandwiches) will be served during the last session

14:00 end of the workshop

We acknowledge the Netherlands Organisation for Scientific Research for financial support.

Eleni Aristodemou (University of Amsterdam)
Semiparametric identification in panel data discrete response models

Abstract: This paper studies semiparametric identification in linear index discrete response panel data models with fixed effects. Departing from the classic binary response static panel data model, this paper examines identification in the binary response dynamic panel data model and the ordered response static panel data model. It is shown that under mild distributional assumptions on the fixed effect and the time-varying unobservables, point-identification fails but informative bounds on the regression coefficients can still be derived. Partial identification is achieved by eliminating the fixed effect and discovering features of the distribution of the unobservable time-varying components that do not depend on the unobserved heterogeneity. Numerical analyses illustrate how the outer bounds change as the support of the explanatory variables varies.

Joerg Breitung (University of Cologne)
Estimation of heterogeneous panels with systematic slope variations (with Nazarii Salish)

Abstract: We analyse estimation procedures for the panel data models with heterogeneous slopes. Specifically we take into account a possible dependence between regressors and heterogeneous slope coefficients, which is referred to as systematic variation. It is shown that under relevant forms of systematic slope variations (i) the pooled OLS estimator is severely biased, (ii) Swamy's GLS estimator is inconsistent if the number of time periods T is fixed, whereas (iii) the mean-group estimator always provides consistent estimators at the risk of high variances. Following Mundlak (1978) we propose an augmentated regression which results in a simple and robust version of the pooled estimator. The latter approach avoids the risk of large standard errors of the mean-group estimator, whenever T is small. We also propose two test statistics for systematic slope variation using the Lagrange multiplier and Hausman principles. We derive their asymptotic properties and provide a local power analysis of both test statistics. Monte Carlo experiments corroborate our theoretical findings and show that for all combinations of N and T the Mundlak-type GLS estimator outperform all other estimators.

Luis Candelaria (University of Warwick)
A semiparametric network formation model with multiple linear fixed effects

Abstract: This paper analyzes a semiparametric model of network formation in the presence of multiple, unobserved, and agent-specific fixed effects. Given agents' observed attributes, the conditional distributions of these effects, as well as the disturbance terms associated with each linking decision are not parametrically specified. I give sufficient conditions for point identification of the coefficients on the observed covariates. This result relies on the existence of at least one continuous covariate with unbounded support. I provide partial identification results when all covariates have bounded support. Specifically, I derive bounds for each component of the vector of parameters when all the covariates have discrete support. I propose a semiparametric estimator for the vector of coefficients that is consistent and asymptotically normal as the number of individuals in the network increases. Monte

Carlo experiments demonstrate that the estimator performs well in finite samples. Finally, in an empirical study, I analyze the determinants of a friendship network using the Add Health dataset.

Andreas Dzemski (University of Gothenburg)
Confidence set for group membership (with Ryo Okui)

Abstract: We develop new procedures to quantify the statistical uncertainty of data-driven clustering algorithms. In our panel setting, each unit belongs to one of a finite number of latent groups with group-specific regression curves. We propose methods for computing unit-wise and joint confidence sets for group membership. The unit-wise sets give possible group memberships for a given unit and the joint sets give possible vectors of group memberships for all units. We also propose an algorithm that can improve the power of our procedures by detecting units that are easy to classify. The confidence sets invert a test for group membership that is based on a characterization of the true group memberships by a system of moment inequalities. To construct the joint confidence, we solve a high-dimensional testing problem that tests group membership simultaneously for all units. We justify this procedure under $N,T\to\infty$ asymptotics where we allow T to be much smaller than N. As part of our theoretical arguments, we develop new simultaneous anti-concentration inequalities for the MAX and the QLR statistics. Monte Carlo results indicate that our confidence sets have adequate coverage and are informative. We illustrate the practical relevance of our confidence sets in two applications.

Gerdie Everaert (Ghent University)
Testing the rank condition for CCE-type estimators (with Ignace De Vos and Vasilis Sarafidis)

Abstract: This paper considers the Common Correlated Effects (CCE) estimation approach of Pesaran (2006) for panels with a multifactor error structure. An important assumption for this methodology to be able to mop up the unobserved common factors is that the matrix of factor loadings is of full row rank. This rank condition translates into the requirement of having at least as many observables holding linearly independent information on the factors as there are factors in the model. This condition is critical for the properties of CCE yet taken for granted in practice. We propose a straightforward method to test the validity of the rank condition for a general class of static CCE-type estimators.

Koen Jochmans (University of Cambridge)
Testing for correlation in error-component models

Abstract: This paper concerns linear models for grouped data with group-specific effects. We construct a test for the null of no within-group correlation beyond that induced by the group-specific effect. The approach tests against correlation of arbitrary form while allowing for (conditional) heteroskedasticity. Our setup covers models with exogenous, predetermined, or endogenous regressors. We provide theoretical results on size and power under asymptotics where the number of groups grows but their size is held fixed. In simulation experiments we find good size control and high power in a wide range of designs. We also find that our test is more powerful than the popular test developed by Arellano and Bond (1991), which uses only a subset of the information used by our procedure.

Geert Mesters (Universitat Pompeu Fabra)
Nonlinear dynamic factor models with interacting level and volatility (with Siem Jan Koopman and Bernd Schwaab)

Abstract: Volatility is an important ingredient in economic and financial decision making and yet the interaction between the levels and volatilities of macroeconomic and financial variables is not well understood. We propose a class of nonlinear dynamic factor models that has factor structures for both levels and volatilities. Both sets of latent factors are modeled jointly in an unrestricted vector autoregressive model. We develop a computationally convenient approximate filtering method for the estimation of all factors. The algorithm relies on numerical integration and can be implemented by augmenting the Kalman filter with weighted least squares regressions. The deterministic model parameters can be estimated by maximum likelihood. Some theoretical bounds and a simulation study show that the methodology is highly accurate when compared to feasible alternative methods. The model is applied in two empirical studies. First, we consider euro area government bond yields between 2008 and 2012 and show that the volatility factor became an economically significant predictor of the yield levels in several countries. Bond purchases by the European Central Bank reduced yields but not the dispersion of pricing errors. Second, the model is applied for forecasting the levels of U.S. macroeconomic variables. We show that the inclusion of interacting volatility factors improves out-of-sample forecasts.

Chris Muris (University of Bristol)
Binarization for panel models with fixed effects

Abstract: In nonlinear panel models with fixed effects and fixed-T, the incidental parameter problem poses identification difficulties for structural parameters and partial effects. Existing solutions are model-specific, likelihood-based, impose time homogeneity, or restrict the distribution of unobserved heterogeneity. We provide new identification results for the structural function and for partial effects in a large class of Fixed Effects Linear Transformation (FELT) models with unknown, time-varying, weakly monotone transformation functions. Our results accommodate continuous and discrete outcomes and covariates, require only two time periods, and impose no parametric distributional assumptions. First, we provide a systematic solution to the incidental parameter problem in FELT. Second, we identify the distribution of counterfactual outcomes and a menu of time-varying partial effects without any assumptions on the distribution of unobserved heterogeneity. Third, we obtain new results for nonlinear difference-in-differences that accommodate both discrete and censored outcomes, and for FELT with random coefficients. Finally, we propose rank- and likelihood-based estimators that achieve \sqrt{n} rate of convergence.

Lorenzo Trapani (Nottingham University)
Determining the dimension of factor structures in non-stationary large datasets (with M. Barigozzi)

Abstract: We propose a procedure to determine the dimension of the common factor space in a large, possibly non-stationary, dataset. Our procedure is designed to determine whether there are (and how many) common factors (i) with linear trends, (ii) with stochastic trends, (iii) with no trends, i.e. stationary. Under the assumption of stationary idiosyncratic errors, this procedure can be used as a cointegration test for a large dataset. Our analysis is based on the fact that the largest eigenvalues of a suitably scaled covariance matrix of the data (corresponding to the common factor part) diverge, as the dimension N of the dataset diverges, whilst the others stay bounded. Therefore, we propose a class of randomised test statistics for the null that the p-th eigenvalue diverges, based directly on the estimated eigenvalue. The tests only requires minimal assumptions on the data, and no restrictions on the relative rates of divergence of N and T are imposed.

Frank Windmeijer (University of Bristol)
Weak instruments, first-stage heteroskedasticity and the robust F-test

Abstract: This paper is concerned with the findings related to the robust first-stage F-statistic in the Monte Carlo analysis of Andrews (2018), who found in a heteroskedastic design that even for very large values of the robust F-statistic, the standard 2SLS confidence intervals had large coverage distortions. This finding appears to discredit the robust F-statistic as a test for underidentification. However, it is shown here that large values of the robust F-statistic do imply that there is first-stage information, but this may not be utilised well by the 2SLS estimator, or the standard GMM estimator. An estimator that corrects for this is a robust GMM estimator, with the robust weight matrix not based on the structural residuals, but on the first-stage residuals. For the grouped data setting of Andrews (2018), this estimator gives the weights to the group specific estimators according to the group specific concentration parameters in the same way as 2SLS does under homoskedasticity, which is formally shown using weak instrument asymptotics. This estimator is much better behaved than the 2SLS estimator in this design, behaving well in terms of relative bias and Wald test size distortion at more standard values of the robust F-statistic. We show that the same patterns can occur in a dynamic panel data model. We further derive the conditions under which the Stock and Yogo (2005) weak instruments critical values apply to the robust F-statistic in relation to the behaviour of this GMM estimator.