Panel and Pooled Data Analysis

PHD COURSE, SAMPOL904

Dr. Vera E. Troeger

August 15 - 19, 2022

UNIVERSITY OF BERGEN

Course Content:

This course will cover various questions and specification issues in multivariate analysis of pooled and panel data and is designed for students who already have training in basic statistics and knowledge of bivariate linear regression analysis. The course deals with different problems arising in panel data analysis when multiple violations of the basic regression assumptions occur because of the nested structure of the data. We will start by quickly discussing the basic Gauss-Markov assumptions of OLS regression analysis, their violations and suitable solutions to such misspecifications, especially when they occur in conjunction. Thus, participants will learn how to deal with different types of heteroskedasticity, spatial correlation, serial correlation and dynamics as well as various kinds of heterogeneity. This discussion will include working with divers data such as cross-sectional, time-series, panel and pooled data. The course gives an overview of the problems arising from complex data structures and also provides techniques to control and account for specific complications. We will also look at problems arising from non-linear relationships, interactions effects and parameter instability. In addition, this course shows how to deal with specification problems such as complex error structures, different kinds of heterogeneity (e.g. unit and slope), dynamic specification issues, missing data, spatial heterogeneity and dependency. Furthermore, we will look at different data generating processes and adequate estimation procedures for e.g. binary choice and limited dependent variable models. Specifically, we will consider truncated and censored data as well as sample selection, instrumental variable approaches and seemingly unrelated as well as simultaneous equation models. The course combines a more theoretical introduction into different topics with practical analysis of diverse data sets using STATA. Students are encouraged to bring their own data sets and present their research projects and empirical analysis during the course.

Course Objectives:

The course requires basic knowledge of inferential statistics, calculus and linear algebra and is designed to further develop the understanding of statistical problems arising from complex data generating processes in applied data analysis. The course mostly deals with questions of specification and model choice and is therefore a very practical course which should enable students to link their empirical models closer to their theoretical arguments and make model choices that are adequate for the data structure at hand. The taught material should help participants to solve their own estimation problems and increase the reliability and efficiency of statistical results. The course is targeted at social and political scientists as well as economists with average statistical skills with a strong interest in applied empirical research and data analysis. The focus lies on practical problems of applied data analysis.

Course Prerequisites:

The course requires average skills and knowledge in inferential statistics, including basic understanding of maximum likelihood and generalized linear estimation methods. In addition, participants should have a basic understanding of matrix algebra and calculus, though the main focus of the course is applied. In addition, participants are required to be familiar with STATA and its command structure. It would be beneficial if participants could write their own do- files and have some familiarity with STATA’s programming language. The course is designed to build on a good working knowledge of cross-section bivariate regression models and basic time-series models. Participants should be able to interpret regression coefficients, standard errors and significance tests.

Reading list

Core Reading:

Wooldridge, Jeffrey M. 2003: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge

Representative Background Reading / References:

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western. Or later edition

Dougherty, Christopher 2002: Introduction to Econometrics, Second Edition, Oxford University Press.

Angrist, Joshua D. and Joern-Steffen Pischke 2009: Mostly Harmless Econometrics, Princeton University Press.

Cameron, Colin A. and Pravin K. Trivedi 2009: Microeconometrics Using Stata, Stata Press.

Gould, William; Pitblado, Jeffrey and William Sribney 2006: Maximum Likelihood Estimation with Stata, Third Edition, Stata Press.

Long, Scott J. and Jeremy Freese 2006: Regression Models for Categorical Dependent Variables using Stata, Second Edition, Stata Press.

Baltagi, Badi H. 2001: Econometric Analysis of Panel Data, Wiley and Sons Ltd.

Hsiao, Cheng, 2003: Analysis of Panel Data, Cambridge University Press, Cambridge.

Data set:

Based on: Garrett, Geoffrey and Deborah Mittchell 2001: “Globalization, Government Spending and Taxation in the OECD.”, *European Journal of Political Research* 39, 145-177.

Course Structure

Day 1:

1. Introduction to the Course and Quick Recap: the OLS model, Gauss-Markov Assumptions and Violations
2. Heteroskedasticity, cross-sectional correlation, multicollinearity, omitted variable bias: tests and common solutions, specification issues 1: Non-Linearities and Interaction Effects

Day 2:

3. Dynamics, serial correlation and dependence over time
4. Pooling Data 1: Heterogeneity: How to choose the right model – how good are the available tests.

Day 3:

5. Pooling Data 2: Slope and Parameter Heterogeneity: Seemingly Unrelated Regression and Random Coefficient Models, modelling parameter instability
6. Endogeneity and spatial econometrics: instrumental variable approaches and simultaneous equation models

Day 4:

7. Limited Dependent Variable Models I: Binary and Ordered Choice Models
8. Limited Dependent Variable Models II: Models for Truncated, Censored and Selected data.

Day 5:

9. Specification Issues in Limited Dependent Variable Models: Dynamics and Heterogeneity
10. Wrap Up, Q&A, Student Presentations

Indicative Readings:

Day 1:

Wooldridge, Jeffrey M. 2003: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, chpt. 4, pp. 49-83.

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpts. 2-3, pp. 22-113.

Dougherty, Christopher 2002: Introduction to Econometrics, Second Edition, Oxford University Press, chpts. 2-4, pp. 48-149.

Wooldridge, Jeffrey M. 2003: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, chpt. 4, pp. 49-83.

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpts. 6-9, pp. 182-317.

Beck, Nathaniel and Jonathan Katz 1995: What to do (and not to do) with Time-Series Cross-Section Data, American Political Science Review 89: 634-647.

Wilson, Sven E. and Danial M. Butler 2007: A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications: Political Analysis 15: 101-123.

Brambor, Thomas, Clark, William Roberts and Matt Golder 2006: “Understanding Interaction Models: Improving Empirical Analysis”, *Political Analysis* 14, 63-82.

Pluemper, Thomas and Christian W. Martin 2003: “Democracy, Government Spending, and Economic Growth: A Political-Economic Explanation of the Barro-Effect.”, *Public Choice* 117, 27-50.

Day 2:

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpts. 10 -12, 18, pp. 323-424, 600-645.

Williams, Laron K. and Guy D. Whitten 2011: “Dynamic Simulations of Autoregressive Relationships.” *The Stata Journal* 11, 577-588.

Keele, Luke and Suzanna De Boef 2004: “Not Just for Cointegration: Error Correction Models with Stationary Data”, unpublished manuscript.

Grant, Taylor and Matthew J. Lebo 2014: “You’re Going to Need a Bigger Model: The Single Equation Error Correction Model with Political Time Series”, wp, Stony Brook University.

Adolph, Christopher, Butler, Daniel M. and Sven E. Wilson 2005: “Like Shoes and Shirt, One Size Does Not Fit All: Evidence on Time Series Cross-Section Estimators and Specifications from Monte Carlo Experiments”, unpubl. Manuscript.

Wawro, Gregory 2002: “Estimating Dynamic Panel Data Models in Political Science.” *Political Analysis* 10, 25-48.

Wooldridge, Jeffrey M. 2002: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, chpts. 10-11.

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpts. 13 -14, pp. 425-484.

Baltagi, Badi H. 2001: Econometric Analysis of Panel Data, Wiley and Sons Ltd., chpts. 2-4.

Hsiao, Cheng, 2003: Analysis of Panel Data, Cambridge University Press, Cambridge, chpts. 3-4.

Plümper, Thomas, Troeger, Vera E. and Philip Manow 2005: “Panel Data Analysis in Comparative Politics. Linking Method to Theory.” *European Journal of Political Research* 44, 327-354.

Beck, Nathaniel 2001: “Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years?” *Annual Review of Political Science* 4, 271-293.

Plümper, Thomas and Vera E. Troeger 2007: “Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects.” *Political Analysis* 15, 124-139.

Plümper, Thomas and Vera E. Troeger 2011: “Fixed Effects Vector Decomposition: Properties, Reliability and Instruments.” *Political Analysis* 19, 147-164.

Beck, Nathaniel and Jonathan N. Katz 2001: “Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon.” *International Organization* 55, 487-495.

Day 3:

Hsiao, Cheng, 2003: Analysis of Panel Data, Cambridge University Press, Cambridge, chpt. 6.

Baltagi, Badi H. 2001: Econometric Analysis of Panel Data, Wiley and Sons Ltd., chpt. 6.

Beck, Nathaniel and Jonathan N. Katz 2007: “Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments.” *Political Analysis* 15, 182-195.

Plümper, Thomas, Troeger, Vera E. and Philip Manow 2005: “Panel Data Analysis in Comparative Politics. Linking Method to Theory.” *European Journal of Political Research* 44, 327-354.

Wilson, Sven E. and Danial M. Butler 2007: “A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications.” *Political Analysis* 15, 101-123.

Wooldridge, Jeffrey M. 2003: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, chpts. 5-9, pp. 83-247.

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpts. 15 -16, pp. 484-553.

Angrist, Joshua D. and Joern-Steffen Pischke 2009: Mostly Harmless Econometrics, Princeton University Press, chpt. 4, 113-221.

Franzese, Robert J. and Jude C. Hays 2008: Empirical Models of Spatial Interdependence, in Box-Steffensmeier, J. / H. Brady and D. Collier (eds.): Oxford Handbook of Political Methodology, forthcoming.

Franzese, Robert J. and Jude C. Hays 2007: “Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data.” *Political Analysis* 15, 140-164.

Pluemper, Thomas and Vera E. Troeger 2011: “Tax Competition and Income Inequality: Why did the Welfare State Survive?” unpubl. Manuscript.

Plümper, T. and E. Neumayer. 2010. "Model Specification in the Analysis of Spatial Dependence." European Journal of Political Research 49: 418-42.

Neumayer, E., and T. Pluemper. 2010. "Spatial Effects in Dyadic Data." International Organization 64: 145-66.

Neumayer, E., and T. Pluemper. 2010. "Making Spatial Analysis Operational: Commands for Generating Spatial-Effect Variables in Monadic and Dyadic Data." Stata Journal 10: 585-605.

Pluemper, T. and E, Neymayer. 2014. “W” Political Science Research and Methods, forthcoming.

Day 4:

Wooldridge, Jeffrey M. 2003: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, chpt. 15, pp. 451-517.

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpt. 17, pp. 553-564.

King, Gary, and Langche Zeng 2001: “Logistic Regression in Rare Events Data.” *Political Analysis* 9, 137-163.

Wooldridge, Jeffrey M. 2003: Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, chpts. 16,17,19, pp. 517-602, 645-684.

Wooldridge, Jeffrey M. 2003: Introductory Econometrics, South-Western, chpt. 17, pp. 565-599.

Day 5:

Ai, Chunrong and Edward. C. Norton 2003: “Interaction Terms in Logit and Probit Models.” *Economics Letters* 80, 123-129.

Honore Bo E. and Ekaterini Kyriazidou 2000: “Panel Data Discrete Choice models with Lagged Dependent Variables.” *Econometrica* 68, 839-874.

Beck, Nathaniel, Katz, Jonathan N. and Richard Tucker 1998: “Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable.” *American Journal of Political Science* 42, 1260-1288.

Carter, David B. and Curtis S. Signorino 2006: “Back to the Future: Modeling Time Dependence in Binary Data.” Unpubl. Manuscript.

Franzese, Robert J. Jr. and Jude C. Hays 2009: “The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation, and Presentation.” Unpubl. Manuscript.


UNIVERSITAS
BERGENSIS