Limperg Course on Archival Data Analysis

David Veenman (d.veenman@uva.nl)
Amsterdam Business School – University of Amsterdam

October 9–15, 2024

1. Course objective and motivation

The purpose of this course is to provide participants a better conceptual and practical understanding of a wide range of econometric tools and procedures that are relevant to accounting research. The course should be of interest to anyone who intends to apply quantitative research methods in a broad range of accounting research (financial, managerial, auditing, ESE/sustainability, etc.).

After successfully completing introductory econometrics courses, most academics often need more practical knowledge and experience to be able to apply the concepts to accounting research settings. For example, although many well-trained academics are familiar with the theoretical notion of "endogeneity" and are aware of the econometric methods that are available to deal with this issue, such knowledge is typically helpful only when we truly understand the nature of our data, the underlying problem that we aim to solve, and the boundaries of the methods in addressing the problem.

Using a variety of recent academic papers and example datasets in Stata, this course is designed to help participants better (i) appreciate the relevance of textbook econometric methods for accounting research, (ii) understand how to apply these methods in accounting research, and, perhaps most importantly, (iii) understand and recognize the *limitations* of these methods. Along the way, we will also explore the benefits of simulation analyses, which often help to make the common econometric problems, and the potential solutions, much easier to understand than the math.

2. Broad overview of course content and topics

Day	Date	Topic(s)
1 (Wed)	Oct. 9	Endogeneity and controlling for confounding factors
2 (Thu)	Oct. 10	Matching, instrumental variables, and selection models
3 (Fri)	Oct. 11	Fixed effects and difference-in-differences designs
4 (Mon)	Oct. 14	Standard errors
5 (Tue)	Oct. 15	Regression discontinuity design, scaling, and outliers

3. Course setup

This is an intensive and interactive. I typically introduce the main problems and methods using the relevant literature, datasets, and simulations. Participants then present assigned papers from the reading list (the assignment of papers takes place a few weeks before the course starts). To facilitate the discussion and learning experience, all participants are expected to have a recent version of Stata installed on their laptops when participating in the course. A separate guide to using Stata for empirical accounting

research can be found here: https://github.com/dveenman/stataguide.

There are no formal entry requirements for this course in terms of background knowledge. However, participants are highly recommended to have passed at least one introductory econometrics course and to have an understanding of the different types of archival data commonly encountered in quantitative accounting research. Without this background knowledge, the learning experience may be less effective. Basic prior experience with Stata and do-files would be helpful as well, but is not required.

As part of the new guidelines implemented by the Limperg Institute in 2019, the course week itself will form "Part B" of the course. "Part A" consists of a pre-course assignment that should be submitted before the start of the course. All participants are expected to complete this assignment in order to satisfy the entry requirement for participation in the course week. Please see the separate assignment document for more information on this assignment and its deadline.

All sessions will be held on campus at the Amsterdam Business School of the University of Amsterdam, Plantage Muidergracht 12. Each day, I plan to start at 9:30 and finish around 17:30. However, note that these are rough indications of the time needed and we might occasionally need more (or less) time, so please ensure you are available from 9:00–18:00 every day. When you are not available to participate on one of the days, you cannot participate in the course.

4. Assessment

Successful completion of this course is determined by both a sufficient assessment of the pre-course assignment ("Part A") and successful participation in the course week itself ("Part B"). More formally, the assessment will be structured as follows:

1. Pre-course assignment: 40% (must be sufficient for entry into course)
2. Assigned paper presentation(s): 30%
3. Participation and contribution to class discussion: 30%

All participants are expected to have carefully read *all* required material *before* the start of the course. Also note that the course assessment will not be based on participants' end-level of knowledge and skills, but rather based on the demonstration of effort in enhancing their knowledge and skills.

5. Setup of paper presentations

Presentations should be prepared to last approximately 20 minutes without interruption (maximum of 20 slides). To facilitate discussion, however, each presenter will have a time slot of about 60 minutes. All other participants are expected to prepare and ask relevant questions during the presentation. Please note that clarification questions are also relevant questions!

Because these papers were chosen for their use or discussion of a specific research method, the emphasis of the presentation should be on the empirical part of the paper. Of course, it is also important to understand the setting and concepts examined in a specific paper, but a rich description of the prior related literature or institutional setting may be less relevant for this course and should therefore receive a bit less weight in the presentation (please email me if you prefer additional guidance).

6. Detailed overview of topics and study material

Below you can find the details on the required reading material for each day and topic. A selection of papers, marked by an asterisk (*), will be assigned to participants for presentation a few weeks before the start of the course.

Day 1: Endogeneity and controlling for confounding factors

We start with a general introduction and review of the fundamental assumptions underlying the estimation of linear regression models using ordinary least squares (OLS), since this is the primary method used in accounting research. Following this discussion we will focus on the reasons, and methods available, to "control for" confounding factors in accounting research settings. Doing so, we will also discuss and illustrate the difference between observable and unobservable confounding factors.

Required reading material:

- Roberts and Whited [2012]: Chapters 1-2
- Gow, Larcker, and Reiss [2016]: Sections 1–3
- Chenhall and Moers [2007]
- * Whited, Swanquist, Shipman, and Moon [2022]
- * Chen, Hribar, and Melessa [2018]
- Leung and Veenman [2018]: Sections 4.2 and 5.2 only

Day 2: Matching, instrumental variables, and selection models

We continue our discussion on controlling for confounding factors by reviewing the "potential outcomes" framework for causal inference. We do so by moving the discussion to methods that control for confounding factors based on matching (propensity-score matching and entropy balancing) and discuss their benefits and drawbacks. We will also focus on the difference between, and complementarities of, matching and regression. In the second part of the day, we focus on the use of instrumental variables and selection models in accounting research.

Required reading material (matching):

- Roberts and Whited [2012]: Chapter 6
- * Armstrong, Ittner, and Larcker [2012]
- DeFond, Erkens, and Zhang [2016]

Leung and Veenman [2018]: Sections 4.2 and 5.2 only

* Shipman, Swanquist, and Whited [2017]

* McMullin and Schonberger [2022]

Required reading material (instrumental variables and selection models):

Roberts and Whited [2012]: Chapter 3

* Larcker and Rusticus [2010]

Lennox, Francis, and Wang [2012]

Jiang [2017]

Donelson, Kettell, McInnis, and Toynbee [2022]

Day 3: Fixed effects and difference-in-differences estimation

We continue our discussion on confounding effects by focusing on fixed effects. The benefit of fixed effects is that they can control for unobservable confounding effects when these effects are constant within a specified group of observations. The discussion of this topic will be illustrated by simulations and a hands-on exercise to replicate some of the key results of Christensen, Hail, and Leuz [2013]. We next move to the use of difference-in-differences estimations as another tool that can be used to control for confounding effects, and which is often combined with fixed effects. We also focus on the benefits and drawbacks of a special type of difference-in-differences design called "staggered" difference-in-differences, which exploits variation in shocks across firms (or groups of firms) and over time.

Required reading material (fixed effects):

Roberts and Whited [2012]: Chapter 7

Christensen et al. [2013]

Gormley and Matsa [2014]

* Breuer and Dehaan [2024]

* Jennings, Kim, Lee, and Taylor [2024]

Required reading material (difference-in-differences):

Roberts and Whited [2012]: Chapter 4

Baruffaldi, Simeth, and Wehrheim [2024]

* Armstrong, Glaeser, and Huang [2022]

* Baker, Larcker, and Wang [2022]

Day 4: Standard errors

While the previous sessions focused on the violation of the exogeneity assumption underlying OLS, which explains the need to control for confounding factors, we now focus on an OLS assumption that relates to the distribution and independence of regression

errors. Understanding this assumption and its violations is essential for the calculation of the precision of our regression estimates. The focus will be on the understanding and computation of cluster-robust standard errors, but we will also discuss the usefulness and application of bootstrapping methods for the estimation of standard errors.

Required reading material:

Bertrand, Duflo, and Mullainathan [2004]

Petersen [2009]

* Gow, Ormazabal, and Taylor [2010]

* Chen, Hribar, and Melessa [2023]

Additional useful reading material (*not required*):

Cameron, Gelbach, and Miller [2008]

Cameron, Gelbach, and Miller [2011]

Cameron and Miller [2015]

Conley, Gonçalves, and Hansen [2018]

MacKinnon, Nielsen, and Webb [2023]

Day 5: Regression discontinuity design, scaling, and outliers

On the final day, we will discuss several more specific topics. A regression discontinuity design (RDD) is a special form of an instrumental variable design. We will discuss how RDDs are implemented in accounting research and focus on their intricacies and pitfalls. We then move on to a discussion about the importance and consequences of scaling accounting variables by measures of firm "size". We finish with an evaluation of the need and methods available to treat outliers in archival research, which brings us back to the fundamental assumptions underlying OLS and highlights the importance of understanding the nature and distributional properties of our data.

Required reading material (regression discontinuity design):

Roberts and Whited [2012]: Chapter 5

Iliev [2010]

* Tan [2013]

Required reading material (scaling and outliers):

Easton [1998]

* Leone, Minutti-Meza, and Wasley [2019]

* Gassen and Veenman [2024]

References

Armstrong, C. S., S. A. Glaeser, and S. Huang. 2022. Contracting with Controllable Risk. *The Accounting Review* 97 (4): 27–50.

Armstrong, C. S., C. D. Ittner, and D. F. Larcker. 2012. Corporate Governance, Compensation Consultants, and CEO Pay Levels. *Review of Accounting Studies* 17 (2): 322–351.

Baker, A. C., D. F. Larcker, and C. C. Y. Wang. 2022. How Much Should We Trust Staggered Difference-in-Differences Estimates? *Journal of Financial Economics* 144 (2): 370–395.

Baruffaldi, S., M. Simeth, and D. Wehrheim. 2024. Asymmetric Information and R&D Disclosure: Evidence from Scientific Publications. *Management Science* 70 (2): 1052–1069.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics* 119 (1): 249–275.

Breuer, M. and E. Dehaan. 2024. Using and Interpreting Fixed Effects Models. *Journal of Accounting Research* 62 (4): 1183–1226.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics* 90 (3): 414–427.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics* 29 (2): 238–249.

Cameron, A. C. and D. L. Miller. 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources* 50 (2): 317–372.

Chen, W., P. Hribar, and S. Melessa. 2018. Incorrect Inferences When Using Residuals as Dependent Variables. *Journal of Accounting Research* 56 (3): 751–796.

Chen, W., P. Hribar, and S. Melessa. 2023. Standard Error Biases When Using Generated Regressors in Accounting Research. *Journal of Accounting Research* 61 (2): 531–569.

Chenhall, R. H. and F. Moers. 2007. The Issue of Endogeneity within Theory-Based, Quantitative Management Accounting Research. *European Accounting Review* 16 (1): 173–196.

Christensen, H. B., L. Hail, and C. Leuz. 2013. Mandatory IFRS Reporting and Changes in Enforcement. *Journal of Accounting and Economics* 56 (2–3, Supplement 1): 147–177.

Conley, T., S. Gonçalves, and C. Hansen. 2018. Inference with Dependent Data in Accounting and Finance Applications. *Journal of Accounting Research* 56 (4): 1139–1203.

DeFond, M., D. H. Erkens, and J. Zhang. 2016. Do Client Characteristics Really Drive the Big N Audit Quality Effect? New Evidence from Propensity Score Matching. *Management Science* 63 (11): 3628–3649.

Donelson, D. C., L. Kettell, J. McInnis, and S. Toynbee. 2022. The Need to Validate Exogenous Shocks: Shareholder Derivative Litigation, Universal Demand Laws and Firm Behavior. *Journal of Accounting and Economics* 73 (1): 101427.

Easton, P. D. 1998. Discussion of Revalued Financial, Tangible, and Intangible Assets: Association with Share Prices and Non-Market-Based Value Estimates. *Journal of Accounting Research* 36: 235–247.

Gassen, J. and D. Veenman. 2024. Outliers and Robust Inference in Archival Accounting Research. *Working paper, Humboldt University Berlin and University of Amsterdam* .

Gormley, T. A. and D. A. Matsa. 2014. Common Errors: How to (and Not to) Control for Unobserved Heterogeneity. *Review of Financial Studies* 27 (2): 617–661.

Gow, I. D., D. F. Larcker, and P. C. Reiss. 2016. Causal Inference in Accounting Research. *Journal of Accounting Research* 54 (2): 477–523.

Gow, I. D., G. Ormazabal, and D. J. Taylor. 2010. Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research. *The Accounting Review* 85 (2): 483.

Iliev, P. 2010. The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices. *The Journal of Finance* 65 (3): 1163–1196.

Jennings, J., J. M. Kim, J. Lee, and D. Taylor. 2024. Measurement Error, Fixed Effects, and False Positives in Accounting Research. *Review of Accounting Studies* 29 (2): 959–995.

Jiang, W. 2017. Have Instrumental Variables Brought Us Closer to the Truth. *The Review of Corporate Finance Studies* 6 (2): 127–140.

Larcker, D. F. and T. O. Rusticus. 2010. On the Use of Instrumental Variables in Accounting Research. *Journal of Accounting and Economics* 49 (3): 186–205.

Lennox, C. S., J. R. Francis, and Z. Wang. 2012. Selection Models in Accounting Research. *The Accounting Review* 87 (2): 589–616.

Leone, A. J., M. Minutti-Meza, and C. E. Wasley. 2019. Influential Observations and Inference in Accounting Research. *The Accounting Review* 94 (6): 337–364.

Leung, E. and D. Veenman. 2018. Non-GAAP Earnings Disclosure in Loss Firms. *Journal of Accounting Research* 56 (4): 1083–1137.

MacKinnon, J. G., M. O. Nielsen, and M. D. Webb. 2023. Cluster-Robust Inference: A Guide to Empirical Practice. *Journal of Econometrics* 232 (2): 272–299.

McMullin, J. and B. Schonberger. 2022. When Good Balance Goes Bad: A Discussion of Common Pitfalls When Using Entropy Balancing. *Journal of Financial Reporting* 7 (1): 167–196.

Petersen, M. A. 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies* 22 (1): 435–480.

Roberts, M. R. and T. M. Whited. 2012. Endogeneity in Empirical Corporate Finance. *Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1748604* .

Shipman, J. E., Q. T. Swanquist, and R. L. Whited. 2017. Propensity Score Matching in Accounting Research. *The Accounting Review* 92 (1): 213–244.

Tan, L. 2013. Creditor Control Rights, State of Nature Verification, and Financial Reporting Conservatism. *Journal of Accounting and Economics* 55 (1): 1–22.

Whited, R. L., Q. T. Swanquist, J. E. Shipman, and J. R. Moon, Jr. 2022. Out of Control: The (Over) Use of Controls in Accounting Research. *The Accounting Review* 97 (3): 395–413.