Limperg Course on Archival Data Analysis

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

October 19–25, 2022

1. Course objective and motivation

The purpose of this course is to provide PhD students and junior academics with a better understanding of the main econometric tools 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, etc.).

After successfully completing basic econometrics courses, many junior 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 junior academics are familiar with the theoretical notion of "endogeneity" and are aware of available econometric methods, such knowledge is typically helpful only when they truly understand the nature of their data and the underlying problem they aim to solve.

Using 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 can help make the common econometric problems and solutions more salient.

2. Broad overview of course content and topics

Day	Date	Topic(s)
1 (Wed)	Oct. 19	Endogeneity and controlling for confounding factors
2 (Thu)	Oct. 20	Difference-in-differences and fixed effects
3 (Fri)	Oct. 21	Matching, instrumental variables, and selection models
4 (Mon)	Oct. 24	Standard errors
5 (Tue)	Oct. 25	Regression discontinuity design, scaling, and outliers

3. Course setup

This is an intensive, interactive, and discussion-based course. Typically, I will introduce the main problems and methods using relevant background literature, datasets, and simulations. After each introduction, participants 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 is provided in the Dropbox folder.

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 basic

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 will most likely 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, in room M4.02. Please note that there will not be an option to follow this course online this year. Each day, I plan to start at 9:00 and finish around 17:00. 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 are not allowed to 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 is less relevant

for this course and should therefore receive less weight in the presentation (please email me if you prefer additional guidance).

6. Detailed overview of topics and required material

Below you can find the details on the required reading material for each day and topic. A selection of papers 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 linear regression estimation using ordinary least squares (OLS), as 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. In doing so, we will also discuss the benefits to researchers of using a predictive validity framework (also known as "Libby boxes") for setting up an empirical study, look at the differences between archival (observational) studies and randomized experiments, and discuss important concepts such as internal vs. external validity.

Required reading material:

- Roberts and Whited [2012]: Chapter 2
- Gow, Larcker, and Reiss [2016]: Sections 1–3
- Chenhall and Moers [2007]
- Whited, Swanquist, Shipman, and Moon [2021]
- Donelson, Mcinnis, and Mergenthaler [2013]
- Chen, Hribar, and Melessa [2018]

Day 2: Difference-in-differences and fixed effects

On the second day we continue our discussion on controlling for confounding factors, but dive a bit deeper into the theory and practice of methods commonly used in the literature. We will start by reviewing the "potential outcomes" framework for causal inference and discuss treatment effects and selection bias. Next, we focus on difference-in-differences designs, fixed effects, and a special form of difference-in-differences design called "staggered" difference-in-differences, which exploits variation in shocks in the cross-section and over time.

Required reading material:

- Roberts and Whited [2012]: Chapters 4+7
- Christensen, Hail, and Leuz [2013]
- Gormley and Matsa [2014]
- deHaan [2021]
- Zhou [2001]

- Jennings, Kim, Lee, and Taylor [2022]
- Baker, Larcker, and Wang [2022]

Day 3: Matching, instrumental variables, and selection models

We next move the discussion to methods that control for confounding factors based on matching (propensity-score matching and entropy balancing), instrumental variables, and selection models in accounting research.

Required reading material (matching):

- Roberts and Whited [2012]: Chapter 6
- Austin [2011]
- Armstrong, Ittner, and Larcker [2012]
- DeFond, Erkens, and Zhang [2016]
- Shipman, Swanquist, and Whited [2016]
- 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 4: Standard errors

While the first days focused on the violation of the exogeneity assumption underlying OLS, which explains the need to control for confounding factors, on this day we will 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 measures of the uncertainty around the parameters we estimate in regressions. 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. Because the proper calculation of standard errors for our coefficient estimates is so important for the inferences we draw from our research, and because standard errors are often considered a difficult and abstract concept, I decided to spend an entire day on this topic.

Required reading material:

- Bertrand, Duflo, and Mullainathan [2004]
- Petersen [2009]
- Gow, Ormazabal, and Taylor [2010]

- Chen, Hribar, and Melessa [2022]
- MacKinnon, Nielsen, and Webb [2022]

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]

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 [2022]

References

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.

Austin, P. C. 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* 46 (3): 399–424.

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.

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.

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. 2022. Incorrect Inferences When Using Generated Regressors in Accounting Research. *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3724730* .

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.

deHaan, E. 2021. Using and Interpreting Fixed Effects Models. *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3699777* .

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.

Donelson, D. C., J. M. Mcinnis, and R. D. Mergenthaler. 2013. Discontinuities and Earnings Management: Evidence from Restatements Related to Securities Litigation. *Contemporary Accounting Research* 30 (1): 242–268.

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. 2022. Outliers and Robust Inference in Archival Accounting Research. *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3880942* .

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. N., J. M. Kim, J. A. Lee, and D. J. Taylor. 2022. Measurement Error, Fixed Effects, and False Positives in Accounting Research.

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.

MacKinnon, J. G., M. Nielsen, and M. D. Webb. 2022. Cluster-Robust Inference: A Guide to Empirical Practice. *Journal of Econometrics, forthcoming* .

McMullin, J. L. and B. Schonberger. 2022. When Good Balance Goes Bad: a Discussion of Common Pitfalls When Using Entropy Balancing. *Journal of Financial Reporting, forthcoming* .

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. 2016. 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. 2021. Out of Control: The (Over)Use of Controls in Accounting Research. *The Accounting Review forthcoming* .

Zhou, X. 2001. Understanding the Determinants of Managerial Ownership and the Link Between Ownership and Performance: Comment. *Journal of Financial Economics* 62 (3): 559–571.