Limperg Course on Archival Data Analysis

David Veenman
University of Amsterdam

October 2020

Course objective and motivation:
The goal of this course is to provide PhD students and junior academics with a better understanding of how to apply key econometric techniques in accounting-research settings. The course should be of interest to those who conduct any type of quantitative research method, regardless of whether the research data used is archival (e.g., financial, managerial, or auditing), experimental, or survey-based.[1]

After successfully completing more general and theoretically-focused econometrics courses, many junior academics often still need more training and experience to apply the general concepts to specific settings in accounting research. For example, although many well-trained junior academics are familiar with the theoretical notion of "endogeneity" and are aware of the available methods, such knowledge is typically not sufficient. The reason is that textbook applications of econometric methods work effectively only when the researcher truly understands the underlying problem and when s/he truly understands the nature of the data. Understanding the nature of the data is something that requires time, practice, and exposure to different data structures common in accounting.

Using recent academic papers and examples of actual accounting-research datasets in Stata, the goal of this course is 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 make use of simulation analyses in Stata, which can help to better visualize the common econometric problems and solutions. We also will allocate some time to discuss best practices in data management, code and data storage, and the importance of transparency in research for the purpose of replicability.

Course setup:
This intensive five-day course will be interactive and discussion-based. The lecturer will introduce each of the main problems and methods using relevant background literature, datasets, and simulations. After each introduction, course participants subsequently 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, to have had basic experience with Stata and do-files, 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.

As part of the new guidelines implemented by the Limperg Institute in 2019, the course week itself forms "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 the deadline.

[1] For a useful and more complete characterization of the different quantitative research approaches, see Figure 3 of Bloomfield, Nelson, and Soltes (2016, "Gathering Data for Archival, Field, Survey, and Experimental Accounting Research", *Journal of Accounting Research* 54(2): 341-395).

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: 30%
2. Assigned paper presentation: 30%
3. Participation and contribution to class discussion: 40%

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 the participant's end-level of knowledge and skills, but rather based on the demonstration of effort in *enhancing* the participant's knowledge and skills.

Setup of paper presentations:
Presentations should be prepared to last approximately 20 minutes without interruption (maximum of about 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 (after having given the presenter about five uninterrupted minutes to introduce the key problem and summarize the primary insights from the paper). 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 the purpose of this course and should therefore receive less weight in the presentation.

Broad overview of course content and topics:

Day	Date	Topics
1	7-Oct	Introduction + Controlling for confounding factors
		Controlling for confounding factors
2	8-Oct	Controlling for confounding factors
		Controlling for confounding factors
3	9-Oct	Matching vs. regression
		Transparency and replication
4	12-Oct	Standard error correction
		Scaling and outliers
5	13-Oct	Instrumental variables and selection models
		Regression discontinuity design (RDD)

Detailed overview of course content and topics:

Day 1: Observational vs. experimental studies, confounding effects, and control variables

Required reading material:

- Angrist and Pischke [2009]: Chapters 1 and 2
- Roberts and Whited [2012]: Chapter 2
- Gow, Larcker, and Reiss [2016]: Sections 1-3
- Donelson, Mcinnis, and Mergenthaler [2013]
- Swanquist and Whited [2018]

Additional useful background reading (not required):

- *Gassen [2014]*
- *Chenhall and Moers [2007]*
- *Chen, Hribar, and Melessa [2018]*

Day 2: More on (non-linear) confounding effects, diff-in-diff, and fixed effects

Required reading material:

- Roberts and Whited [2012]: Chapter 4+7
- Irani and Oesch [2013]
- Gassen, Skaife, and Veenman [2020]: Introduction, Section 5, and Appendix B
- Gormley and Matsa [2014]
- Christensen, Hail, and Leuz [2013]

Additional useful background reading (not required):

- *Angrist and Pischke [2009]: Chapter 3+5*
- *Zhou [2001]*

Day 3: Matching vs. regression + Transparency and replication

Required reading material for matching vs. regression:

- Roberts and Whited [2012]: Chapter 6
- DeFond, Erkens, and Zhang [2016] Comments and response [2017]
- Armstrong, Ittner, and Larcker [2012]
- Shipman, Swanquist, and Whited [2016]
- Leung and Veenman [2018], Sections 4.2 and 5.2

Additional useful background reading (not required):

- *Stuart [2010]*
- *Core [2010]*
- *Lawrence, Minutti-Meza, and Zhang [2011]*

Required material for transparency and replication:

- Ball and Brown [2019], Section 7 only
- Harvey [2019]
- JAR 2019 conference video (first 40 minutes): https://media.chicagobooth.edu/Mediasite6/Play/298cc9782d7444818fe55c1c86bd8e731d

Day 4: Standard error corrections, scaling and outliers

Required reading material:

- Petersen [2009]
- Gow, Ormazabal, and Taylor [2010]
- Easton and Sommers [2003]
- Leone, Minutti-Meza, and Wasley [2019]
- Gassen and Veenman [2020] (note: paper will be updated before the course)

Additional useful background reading (not required):

- *Angrist and Pischke [2009]: Chapter 8*
- *Bertrand, Duflo, and Mullainathan [2004]*

Day 5: Instrumental variables, selection models, and regression discontinuity design

Required reading material:

- Roberts and Whited [2012]: Chapter 3+5
- Lennox, Francis, and Wang [2012]
- Larcker and Rusticus [2010]
- Iliev [2010]
- Tan [2013]

Additional useful background reading (not required):

- *Angrist and Pischke [2009]: Chapter 4+6*
- *Tucker [2010]*
- *Lee and Lemieux [2010]*

Reference list (papers posted in Dropbox folder):

Angrist, J. D., and J.-S. Pischke. "*Mostly Harmless Econometrics: An Empiricist's Companion*" (2009).

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

Ball, R., and P. Brown. "Ball and Brown (1968) After Fifty Years". *Pacific-Basin Finance Journal* 53 (2019): 410–431.

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

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

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

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

Core, J. E. "Discussion of Chief Executive Officer Equity Incentives and Accounting Irregularities". *Journal of Accounting Research* 48 (2010): 273–287.

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

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

Easton, P. D., and G. A. Sommers. "Scale and the Scale Effect in Market-based Accounting Research". *Journal of Business Finance & Accounting* 30 (2003): 25–56.

Gassen, J. "Causal Inference in Empirical Archival Financial Accounting Research". *Accounting, Organizations and Society* 39 (2014): 535–544.

Gassen, J., H. A. Skaife, and D. Veenman. "Illiquidity and the Measurement of Stock Price Synchronicity". *Contemporary Accounting Research*, 37 (2020): 419–456.

Gassen, J., and D. Veenman. " Outliers and Robust Regression in Archival Accounting Research" *Preliminary working paper* (2020).

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

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

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

Harvey, C. R. "Replication in Financial Economics". *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3409466* (2019).

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

Irani, R. M., and D. Oesch. "Monitoring and Corporate Disclosure: Evidence from a Natural Experiment". *Journal of Financial Economics* 109 (2013): 398–418.

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

Lawrence, A., M. Minutti-Meza, and P. Zhang. "Can Big 4 versus Non-Big 4 Differences in Audit-Quality Proxies Be Attributed to Client Characteristics?". *The Accounting Review* 86 (2011): 259–286.

Lee, D. S., and T. Lemieux. "Regression Discontinuity Designs in Economics". *Journal of Economic Literature* 48 (2010): 281–355.

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

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

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

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

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

Shipman, J. E., Q. T. Swanquist, and R. L. Whited. "Propensity Score Matching in Accounting Research". *The Accounting Review* 92 (2016): 213–244.

Stuart, E. A. "Matching Methods for Causal Inference: A Review and a Look Forward". *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25 (2010): 1–21.

Swanquist, Q. T., and R. L. Whited. "Out of Control: The (Over)use of Controls in Accounting Research". *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3209571* (2018).

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

Tucker, J. W. "Selection Bias and Econometric Remedies in Accounting and Finance Research". *Journal of Accounting Literature* (2010).

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