MHR 8781: Non-Experimental Research Methods in Strategic Management
Michael J. Leiblein
Fall 2023

E-mail: leiblein.1@osu.edu
Office: Fisher 848; Hours be appointment.
Phone: (614) 292-0071
Class: To be announced

Introduction to the Course:

This course provides a Ph.D.-level overview of the statistical tools used to study non-experimental phenomena within and between organizations. The course assumes that participants are familiar with the philosophy of science, basic statistics, and probability theory.

Course Objectives:

This course is designed to help participants improve their ability to apply and evaluate research using methodological approaches commonly used in high-quality, non-experimental management research. The course provides an introduction to issues including the assumptions underlying OLS regression, discrete choice analysis, instrumental variables, panel data analysis, and event history analysis. In addition, the course provides examples of well-received papers in the strategic management field applying these models as well as an opportunity to apply these techniques using datasets provided by the professor. Ultimately, the objective is to help you learn to conduct independent empirical research and to critically evaluate others’ empirical research.

The course pays minimal attention to mathematical proofs and other technical matters. I suggest participants consider purchasing textbooks by Bill Greene, Jeff Wooldridge, or others listed in the supplementary readings section of the syllabus to review these matters. While the course offers basic opportunities to work with data, you should also enroll in courses or purchase manuals to help develop your skills with programs such as “R”, SAS, or Stata. Again, recommendations are offered in the supplementary readings section of the syllabus.

Instructional Procedure:

The course will be taught using a seminar style. Each session includes a set of focal readings (to introduce or review a specific statistical methodology), a set of applied readings (to illustrate how others have applied this method), and a set of supplemental readings (to allow you to delve deeper into particular topics). Finally, some sessions will include a short problem set assignment to allow you to “get your hands dirty” manipulating and/or analyzing data.

In contrast to other seminars that cover particular theories or topics, this seminar will aim, individually and collectively, to develop practical responses to basic questions regarding non-experimental research design. I have also included a set of questions to guide your thinking before, during, and after each in-class discussion.

Evaluation:

The grading plan describes the relative importance attached to each of the individual activities used to assign a course grade. The overall course grade will reflect your performance in terms of: (1) In-Class Discussion (35%), (2) Problem Sets (30%), and (3) Empirical Project (35%).

1. In-class discussion (35%). Effective participation in this course involves reading all assignments, writing short responses to assigned study questions prior to class, and participating in all class discussions. Simply reading the assigned materials is insufficient to earn a good class contribution grade. You must be able to contribute to the discussion—often by articulating cogent responses to the discussion questions. Outlines of written responses to weekly discussion questions should be uploaded to

---

1 I have benefited from conversations with Javier Gimeno, Ashton Hawk, Glenn Hoetker, Tammy Madsen, & Jeff Reuer while developing this syllabus.
Carmen Canvas at least 24 hours before the corresponding class discussion.²

2. **Problem Sets (30%).** I have designed several problem sets to help you become familiar with applying the statistical methodologies covered in this course.³ These problem sets provide an opportunity to apply the methodology associated with the readings (and to foster additional in-class discussion). *I encourage you to work with others on these assignments*. You may even turn in assignments representing the work of the group as a whole. Problem set responses should be uploaded to Carmen Canvas at least 24 hours before the corresponding class discussion. You are on your honor to only associate your name with problem sets to which you have contributed.

3. **Empirical Project Proposal (35%).** Research creativity is a fundamental skill required for successful scholarship. As you read the literature, you should always look for opportunities to contribute to the literature. The empirical research project proposal asks you to develop a presentation describing an empirical study that you might complete in the future. While findings are not required, this presentation should state the research question, tested hypotheses, data source, measures, and empirical methodology required to complete the study. The empirical methodology should depict the structure of the desired data matrix, the equations underlying the estimation procedure, and how the results will test the hypotheses.

Each assignment will be graded as follows:
- Incomplete. Assignment either not submitted or totally unacceptable in quality.
- Check-minus. Sub-standard quality.
- Check. Good and acceptable, but not especially brilliant, interesting, or insightful.
- Check-plus. Truly outstanding in some way.

Grades of check-plus are expected to be awarded rarely, if ever. So, don’t worry if you are not getting check pluses. Only worry about not getting checks.

**Software Used:**

As part of this course, you will need to complete data manipulation and statistical analysis assignments. I do not have any preferences about the software that you use for the analysis (e.g., R, SAS, Stata).

² I recommend that the group consider having a single course participant summarize discussions from each session in a way that will be helpful as a reminder/study aid to the rest of the group.

³ This summary should have a clear file title (e.g., Non-Experimental Methods OLS Regression Last Name.docx), repeat the syllabus questions, highlight the main points in the readings and the class discussion addressing these points, and offer some inference regarding the importance of this discussion for the field of strategic management. If the group can organize this effort, I will review and disseminate these summaries to the class during the ensuing session.

³ I have developed additional problem sets that we do not have time to complete during this seven-week seminar.
Session 1

Strategic Management, Generating (Interesting) Hypotheses, and Hypothesis Testing

Conceptual Readings:


2. Wollman, Lauren F. A short video presentation regarding different research paradigms (https://www.chds.us/ed/research-paradigms/).


Supplementary Readings on Hypothesis Testing (not required):


While the first two sessions include a lot of reading, this material should be familiar. The goal is to review the basic concepts and ensure everyone has developed a solid foundation for future empirical work.

I've included additional readings on the philosophy of science, theory development, and research methodology at the conclusion of the syllabus. These readings are NOT required. I include these and similar lists of additional readings for other sessions to provide you with some guidance should you desire to develop deeper insight on topics of personal interest.

This syllabus concludes with a set of additional topics that we do not have time to cover in a 7 week seminar. I'm open to discussing these topics “off line” as we aim to develop additional expertise.
Supplementary Readings on Theory & Hypothesis Development (not required):


- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. University of Chicago Press, 1970 (especially through page 10). *An important statement on the nature of scientific breakthroughs and a challenge to the concept of normal scientific progression as discussed by Karl Popper.*


Session 1 Problem Set  
Due prior to Session 1

Instructions:
Good empirical researchers faithfully report basic descriptive statistics in their work. One of the reasons reporting descriptive statistics is a good practice is because we are investigating samples. Theoretically, these samples are randomly taken from a general population. We want to maintain transparency so other scholars may evaluate, test, and/or replicate our findings on their samples. Moreover, a good description of your data can help you develop insights and understanding regarding any anomalies or oddities in your data.

This problem set asks you to perform some basic data manipulation processes and simple statistical analyses. The problem set is designed to help you familiarize yourself with a statistical program and gain experience generating insights from descriptive statistics, correlations, and differences in sample means.

I do not have any preferences about the software that you use for the analysis; you are free to use whatever software package you deem appropriate. I encourage you to work with others on these problem-set assignments. You may even turn in assignments representing the work of the group as a whole. If you need a basic resource for Stata, I recommend the following: http://www.ats.ucla.edu/stat/stata/.

1. Import the Game Data and Cognition Data Excel datasets from the Carmen/Canvas course site.
   a. The Cognition Data file reports scores from multiple user IDs on several cognitive tests (we only wish to retain Mind Reading, Race to 100, and Cognitive Reflection).
   b. The Game Data file reports the Game Date, Game ID, User ID, Round, Funds, Marginal Cost, R&D Investment, R&D Outcome (New Marginal Cost), Entry Choice, and Predicted Number of Rivals).
   c. Our objective is to test whether Mind Reading, Race to 100, or Cognitive Reflection Tests are correlated with choices to invest or compete in the game.

2. Please merge the two datasets on the common variable, user-id, and create the following new variables.
   a. Actual Number of Rivals: A count of the number of rivals in a given game and round across the entire sample.
   b. Accuracy: Actual Number of Rivals – Predicted Number of Rivals.

3. Create a table of descriptive statistics and correlations (mean, standard deviation, minimum, and maximum values) for each of the following variables. R&D Investment, Entry Choice, and Accuracy.

4. Conduct a t-test of differences in means for R&D investment, Entry Choice, and Accuracy variables across the top and bottom quartiles of the Mind Reading, Race to 100, and Cognitive Reflection Test scores.

5. What do you observe? What inferences do you draw? Do these findings suggest that cognitive tests are correlated with decisions in strategic factor market competition?

6. Copy your code, tables, and observations into an MS Word document and upload them to the appropriate Carmen dropbox at least 24 hours prior to our next seminar.
Session 1 Preparation Questions

1. What are the components of a theory? What are the components of a good theory? Can you use this definition to evaluate a prominent theory in “strategic management” or “international business”?

2. Stinchcombe describes the concept of the “crucial experiment.” What is a crucial experiment? Could you provide an example of a “crucial experiment” that you have seen in the literature? Is Stinchcombe assuming one of the types of research summarized in Wollman’s video (e.g., descriptive, evaluative, prescriptive, exploratory, and predictive research)?

3. What do we mean when we say that X is the “cause” of Y? What are typical challenges of empirically establishing causation in non-experimental research?

4. Please view the taxonomy below. What do the phrases “good strategy”, “good (disciplinary) social science”, and “good empirical identification” mean to you? Is it valuable to conduct research in regions 2 or 3 of the diagram?

![Taxonomy Diagram]

Source: Chaudhuri, Leiblein, & Rezay (2021)
Session 2
OLS Review: Mediators, Moderators, and Interaction Effects

Conceptual Readings (a lot reading this week; but largely review):


Application Readings (focus on the methods section):

Supplementary Readings:


Session 2 Problem Set
Due prior to Session 2

Instructions:
We all want to design research studies that are realistic and unbiased, offer generalizable truths, and precisely control and measure relevant variables. Research designs vary, however, in their ability to address these basic concerns (see Leiblein & Weber slide below). There are significant tradeoffs.

Historically, research in the field of strategic management has emphasized non-experimental research methods. Whereas experimental research allows for control via the experiment itself (e.g., random assignment to a treatment group, same conditions, materials, etc.), non-experimental field studies utilize statistical control. While we cannot rely on random assignment, non-experimental studies use quasi-experimental designs that develop criteria to assign and compare control and treated groups.

The objective of this problem set is to help you familiarize yourself with one of the basic tools to achieve statistical control in non-experimental studies, multivariate regression.

Again, I do not have any preferences about the software that you use for the analysis (e.g., R, SAS, Stata) and I *encourage* you to work with others on these problem-set assignments. You are simply on your honor to associate your name with problem sets on which you have made some contribution.

1. Import the Leiblein MillerJ NCMM Survey Excel dataset from the Carmen / Canvas course site. This file reports data from a NCMM survey developed to explore whether and how elements of organization and managerial choice affect innovation. The idea was to address the fragile findings in a large literature associating firm size with innovation by proposing that differences in managerial choices (e.g., incentive intensity, decision-making process, resource allocation process) across different sized groups of firms (e.g., small-, medium-, and large-sized) explained differences in innovation outcomes across those organizations.

2. Create a table of descriptive statistics and correlations for the following variables in the dataset (Firm Age (Column C), Firm Size (Column D), Innovation Employees to Total (Column F), Provision of Funds formalized (Column AA), centralized innovation selection process (Q210-1), environmental scanning), product innovation, process innovation, radical product innovation, and radical process innovation.

3. Test for differences in means for these variables across the small-, medium-, and large-firm size categories.

4. Estimate regression models to explore whether and how firm size affects the association between innovation expenditures and the number of new products and services. Do differently sized firms vary in the effectiveness of their innovation investments? Why do you think this is or is not the case? Are there important control variables (e.g., CEO compensation) that affect this baseline association?

5. What do you observe? Do you have concerns with the residuals? Are there outlier observations (from the residual plot, DFFITS or DFBETA statistics)? Can you imagine a better model to test claims that resource allocation and organizational policies affect innovation outcomes?

Copy your code, tables, and observations into an MS Word document and upload to the appropriate Carmen dropbox at least 24 hours prior to our next seminar.
Session 2 Preparation Questions

1. The regression model includes an error term. What is that error term supposed to represent? Why do we assume that the error term is distributed as a normal distribution?

2. What are “control variables” and why are they important?
   a. A reviewer observes that your paper does not control for a variable that has a well-known association with your dependent variable. The reviewer asks that you add the variable. However, it may not be feasible to collect good field data on this variable. Under what conditions could you make a methodologically credible argument that the lack of the control variable does not pose a threat to the statistical conclusion validity of your analysis?
   b. Suppose that your theory predicts a positive relationship between X and Y. Yet, you find a negative correlation. You suspect that the difference between theory and correlation is that you are not controlling for other factors that may influence the DV. You have a list of potential control variables, with information about their correlation with X and Y. How would you identify what missing control variables are likely to explain the unexpected results?

3. Should you include the main effects when you test for interaction effects? Why? Why not? What are the pros and cons of including the main effects? In a model with interaction effects, how should we interpret the main effects?

4. What are the main findings in the Schilke and Rodriguez & Nieto papers? Do you believe it would be useful to replicate these studies? Why or why not? Can you devise an extension?
Session 3
Discrete Choice Analysis (& Replication)

Conceptual Readings:


Application Readings (focus on the methods section):


Supplementary discrete choice models readings (not required):

- Maddala, GS. 1983. Limited Dependent and Qualitative Variables in Econometrics. Cambridge. (See especially, chapter 2).

Supplementary discrete choice estimation readings (not required).


Supplementary limited dependent variable model readings (not required; also event history):


Supplementary Replication readings (not required)

Session 3 Problem Set
Due Prior to Session 3

Instructions:
In the social sciences, scholars are often interested in exploring issues involving choices that are measured as dichotomous or nominal dependent variables. This is particularly true in strategic management where our focus is on sets of interdependent choices (e.g., choices to organize via a firm or market; choices to diversity in a related or unrelated fashion; choices to expand in domestic or non-domestic markets; estimates of complementarity across a bundle of resources).

The existence of a discrete dependent variable introduces a series of statistical complications. In particular, residuals tend to be skewed. If residuals are skewed then parameter estimates may be wrong and standard errors will be wrong. As we’ll discuss in this session, the existence of discrete dependent variables also complicates the interpretation of statistical estimates.

The objective of this problem set is to help you familiarize yourself with estimation using logit and probit models. The assignment asks you to manipulate a subset of responses from a paper I published with Doug Miller (a successful OSU alumnus).

Again, I do not have any preferences about the software that you use for the analysis (e.g., R, SAS, Stata) and I *encourage* you to work with others on these problem-set assignments. You are simply on your honor to associate your name with problem sets on which you have made some contribution.

1. Import the Leiblein Miller Make-Buy dataset from the Carmen / Canvas course site. This data was developed to provide an early test of the importance of transaction cost and resource-based arguments on governance choice and whether these choices affected performance outcomes. The Leiblein & Miller paper was published the same year as Ai & Norton’s article on interpreting interaction terms in logistic regression. This led to questions regarding whether additional insight could be developed using the L&M data.

2. Attempt to replicate the results in Leiblein and Miller (2003). It may be helpful to create a table of descriptive statistics and correlations as an initial step.

3. What do you observe? What is the overall fit of your model? What does it mean to report a “pseudo R²” term? How do you interpret your coefficients? Have you replicated our results (or found an error)?

4. Copy your code, tables, and observations into an MS Word document and upload to the appropriate Carmen dropbox at least 24 hours prior to our next seminar.
Session 3

Preparation Questions

1. What’s wrong with estimating an OLS models with a dichotomous (yes or no, stay or go, etc.) or ordinal (low, medium, high) dependent variable?

2. Suppose that the DV is categorical but not binary (e.g., governance choice, country to enter, board structure). How would you choose amongst the use of an ordered or multinomial logit (probit) model? Are there other possible methodologies that you could use to study such a DV? List both the ones discussed in the readings and others that you know about. Can you identify a decision-tree to suggest which method is appropriate for which type of question?

3. As you read in this session, Glenn Hoetker and Bennet Zelner have replicated and extended the Leiblein and Miller (2003) results. More recently, Jordan Siegel and his students borrowed the data used in the Leiblein & Miller paper and published an extended model. If you met Glenn, Bennet, or Jordan at a conference, what question would you ask them?

4. Bring a topic of interest to class involving a discrete choice. After describing this choice, describe the choice set (the set of all possible responses to the DV). Be prepared to discuss how different definitions of the choice set may affect the design of your study.
Session 4
Instrumental Variables

Conceptual Readings:


Application Readings (focus on the methods section):


*The above image summarizes the Bradford Hill criteria used for assessing causality in epidemiology.*
Supplementary Conceptual Readings (not required):


Supplementary Application Readings (not required):

Session 4 Preparation Questions

1. There is an old maxim in statistics that one “should not sample on the dependent variable”. Essentially, this maxim dissuades from going out and picking a group of observations because they exhibit very high or low values of the variable that you are interested in. For instance, if you are interested in firm performance, you should not go out and collect a group of best-performing firms and then draw implications from that sample. What is the statistical logic of that maxim? What threats to validity are involved in the practice of sampling on the DV?

2. Suppose that you conducted a survey of a sample population and obtained a 50% response rate. A reviewer is making a big fuss that your sample may not be representative of the population? What can you do, either in terms of logical argumentation or statistical argumentation, to defend yourself from that claim? What tests could you carry out to show that the sample is not biased?

3. Why are we concerned about sample selection bias? What are some problems (say three) you are aware of from "sample selection bias"?

4. Several papers in this section state that strategy is essentially self-selected (i.e., firms are likely to take whatever strategy is best for them). What are the implications of that view for the analysis of the strategy-performance relationship? (For OB and HR students: human behavior may be goal oriented; what are the implications about the assessment of the behavior-goal relationship?).

5. Please identify and briefly summarize a study in your area of research where you could apply the theory and methods of self-selection.
Session 5: Complementarity

Conceptual Readings:


Application Readings (focus on the methods section):


Supplementary Readings:


Sessions Problem Set
Due Prior to Session 5

Instructions:
In a recent paper, Marcus Larsen, Torben Pedersen, and I have explored whether the choice to organize via a firm or market and the choice to participate in a given country are independent. Our primary claim is that the employment of distinct combinations of governance and location choices suggests that these choices may be interdependent. We note examples where firms’ choice sets appear to consider both governance and location simultaneously and that the existing empirical evidence largely assumes that these choices are independent. 5, 6

The objective of this problem set is to help you consider ways in which one might test for the existence of “interdependence” across choice sets with interdependent choices. For instance, Leiblein, Larsen, and Pedersen estimate simple tests of association (Cochran-Mantel-Haenszel), use a bivariate probit model to jointly estimate multiple choices (two regressions with correlated error terms), and apply 2SLS (two-state least squares regressions) with instrumental variables to address the possibility that any association between governance and location choices are driven by attributes of the choices themselves. I ask you to consider the tradeoffs across these methods and to replicate our CMH and/or bivariate probit results. I also ask that you consider questions where this basic logic that may be relevant to your research.

Again, I do not have any preferences about the software that you use for the analysis (e.g., R, SAS, Stata) and I *encourage* you to work with others on these problem set assignments. You are simply on your honor to only associate your name with problem sets on which you have made some contribution.

1. Import the Leiblein Larsen Pedersen interdependence Buy dataset from the Carmen / Canvas course site. The Leiblein Miller dataset reports

2. Estimate a CMH test and a bivariate probit regression using variables in the dataset that you deem appropriate. In our baseline bivariate probit model, we estimate the following models.
   a. Governance = f{problem complexity, demand uncertainty, complexity*uncertainty, # suppliers, production experience, sourcing experience, firm tenure, and maturity).
      i. Make = f{ass_spec, student_residual, ass_spec*student_residual, exante, lfirmexp, lfirmsource, firmten, maturity}.
   b. Location = f{polcon, wage-level, country tech environ, experience of domestic firms, firm experience in host market, distance, GDP per capita, population, economic growth, and Kogut-Singh}.
      i. Country (e.g., Country as in an agglomeration of Aus, Aut, Bel, Can, etc.) = f{polcon, wage, cte, expferoreign, expferhostloc, gdppericap, population, econgrowth, kogut-singh}.

3. What do you observe? What does the rho parameter imply? How do you interpret your coefficients?

4. Torben Pedersen has completed much of the analysis in our paper. If you met Torben at a conference, what question would you ask him?

Copy your code, tables, and observations into an MS Word document and upload to the appropriate Carmen dropbox at least 24 hours prior to our next seminar.

---

5 We are not the first to make this claim. Many scholars make similar claims (e.g., Buckley, Devinney, and Louviere 2007; Mudambi and Venzin 2010; Teece 2009, 2021; Zhao 2000, etc.) makes similar claims. In part, the issue is driven by data limitations (e.g., Martin, Swaminathan, and Tihanyi, 2007). There are also important methodological challenges (e.g., Nielsen, Asmussen, and Weatherall, 2017). All are worried that empirical work that has not controlled for this potential interdependence may be biased.

6 As a further complication, the location choice literature (which focuses on associations between location parameters such as supply or demand conditions and location choices such as whether to locate in China or France) often uses McFadden’s conditional logit model.
Session 5 Preparation Questions

1. Be ready to discuss the similarities and differences of these concepts: (i) Contingency theory, (ii) Fit, (iii) Moderator relationship, (iv) Interaction effect, and (v) Complementarity.

2. What is fit? What is complementarity? Find an example in your literature of an argument suggesting that two or more variables have a complementary or substitutive relationship. Be prepared to discuss how one might test the existence of such a relationship.
Session 6

Panel Data Analysis (Fixed & Random Effects, Diff n Diff, Event History)

Conceptual Readings:


Application Readings (focus on the methods section):


Supplementary Readings on Count Models (not required):


Supplementary Panel Data Readings (not required):


Supplementary Application Readings (not required):


Supplementary Event History Readings (not required):


Session 6 Preparation Questions

1. What is panel data? Why would anybody want to collect panel data? What are the proposed advantages of panel data? Are some of these advantages particularly helpful for non-experimental and field researchers? What are the costs of panel data?

2. What are the problems of analyzing panel data using the traditional OLS approach – i.e., pooling all observations and running an OLS regression? What are the differences between fixed effects and random effect estimators? What are the pros and cons of each? How would you know if you need to use one or the other? While knowing the formulas is a plus, try to understand the insight behind the formulas.

3. What is event history analysis? What is the difference between censored and truncated samples? What is left-censoring? What is right censoring? What challenges arise (using OLS) in censored or truncated samples? What is the hazard rate? Is it a probability? How does it relate to the survivor function? Can you provide examples of hazard rates and survivor functions relevant to the strategic management literature?

4. What are the different ways to handle time-varying variables? What are the differences between continuous-time and discrete-time methods? What are the advantages/disadvantages of discrete time modeling? What is time (or duration) dependence? Why is it important? What are the ways in which we can model time dependence in an event history model?

5. Find an example in your literature of someone using panel data in their research. Bring a copy of the paper to class and be prepared to discuss why you find this application of panel data techniques interesting—that is how does the paper highlight the advantages of panel data and/or are there any deficiencies or problems associated with the application of panel data in this paper?
Session 7
Preparation Questions

1. Please prepare a detailed presentation of a research project that you will analyze after this seminar, paying particular attention to the hypotheses, research design, and specification of your model. If you have a draft paper, please send me an electronic copy so that I can become more familiar with your project.

   a. Davis depicts theoretical models graphically in systems of boxes and arrows. In your presentation, use a path diagram to represent your model.

   b. Using the ideas discussed by Murray Davis (1971), develop one slide that outlines the introduction of your paper. How are you making your paper interesting? How are you "problematizing" your contribution?

   c. Think about your research project, and about the main relationship that you expect to find. Come up with at least 2 different alternative explanations for that relationship that do not involve your suggested mechanism. How could you control for those alternative theories? How could you build a critical experiment?

   d. Describe the research design you intend to conduct in your project. Why do you believe this design is appropriate?
There are many additional research methodologies that we do not have time to cover in a restricted, seven week seminar. I’ve collected some additional readings that you may find useful to start your own reading in these other areas.

**Doctoral Dissertations in Strategic Management**

- I encourage you to read one or two award winning dissertations early in your doctoral education. This exercise provides you with an opportunity to better understand what a high-quality dissertation looks like before you begin your endeavor.

**Managing the Research Process**

- Starbuck, William H. 1999. Fussy Professor Starbuck’s Cookbook of Handy-Dandy Prescriptions for Ambitious Academic Authors. http://pages.stern.nyu.edu/~wstarbuc/Writing/Fussy.htm
- Discussion Questions:  ♦ How can I make my writing more persuasive? ♦ How should I conduct a thorough and efficient literature review? ♦ How am I supposed to structure a paper so it reads like a journal article? ♦♦ How can I manage my relations with collaborators? ♦♦ Whom should I ask for comments on papers? At what stage? How many people should I ask? How many times can I reasonably ask any one person? ♦♦♦ How does the review process operate? ♦♦♦♦ How should I respond to those $%^&#@!!! reviewers?
Sample Selection and Matching


Research Questions: ♦ What is the goal of matching techniques? ♦ ♦ How do you choose when to use control variables, instrumental variables (including Heckman), 2SLS, or propensity scoring techniques? What can we learn from propensity score matching techniques that we can’t learn from other research designs?

Experiments


Research Questions: ♦ What are the features of good laboratory experiments? ♦ ♦ What are the biggest drawbacks with laboratory experiments?

Qualitative Research


• Research Questions:
  ♦ What is qualitative research?
  ♦♦ What can we learn from qualitative research and case studies that we can’t learn from other research designs? How does research that is primarily inductive differ, in practice, from research that is primarily deductive?
  ♦♦♦ What are the strengths and weaknesses of case studies?

Measurement, Validity, and Reliability (See also Professor Klein’s OBHR methods course).


• Research Questions: How do we determine whether our measures are valid? Whether they have internal validity, construct validity, or external validity? How do we develop reliable and valid measures?
Surveys


- **Research Questions**: ♦ What does it take to construct and conduct good surveys? ♦♦ What are structured interviews and how can I conduct them? ♦♦♦ What are the strengths and weaknesses of these research designs?

Replication


NBER Methods MiniCourse

- Lecture notes and videos on various estimation techniques at http://www.nber.org/minicourse3.html
Instructions:
The introduction of statistical control is a powerful tool. Yet, it too has its limits. Unless the specification of
your model, including the vector of control variables, is perfect, your estimate of associations between
independent and dependent variables will be biased.

Establishing rigorous control is a daunting task. Imagine a researcher wants to better understand how
outsourcing or diversification affects risk or performance outcomes. Theory informs us to explore how
exchange- and firm- effects influence the choice (to diversify) and its consequences. And, we may control for
a broad range of factors (e.g., focal firm age, industrial concentration and technical trajectory in existing and
new product markets, formal and informal institutions in existing and new geographic markets). If you can
confidently claim you accurately measured all the relevant confounding variables, you have a good argument
for conditional independence (of X and Y) and may employ models such as multiple regression. If, however,
after inclusion of your control variables, there is a correlation between the treatment indicator, T, and the
error term, we say that the treatment is ENDOGENOUS. This is problematic since a valid estimate of
treatment effects requires EXOGENOUS regressor variables.

The objective of this problem set is to introduce you to instrumental variable estimation tools that account
for this selection effect. The assignment asks you to manipulate data from a paper I published with Jeff Reuer
and Frederic Dalsace (LRD, 2002) using the Heckman procedure.

Again, I do not have any preferences about the software that you use for the analysis (e.g., R, SAS, Stata) and I
*encourage* you to work with others on these problem set assignments. You are simply on your honor to only
associate your name with problem sets on which you have made some contribution.

1. Import the Leiblein Reuer Dalsace dataset from the Carmen / Canvas course site (upload). The
Leiblein Reuer Dalsace dataset reports performance consequences of individual exchange decisions.

2. Estimate a first stage regression model as per Table 2, Model III in the Leiblein, Reuer, and Dalsace
(LRD, 2002) paper.

3. Estimate the performance consequences of governance decisions using the LRD data (applying 2-
stage method as in LRD or propensity score matching).

4. What do you observe? Do you believe that LRD introduce a “good” instrumental variable? What do
you think LRD mean by “misalignment”? Is there a way to advance this analysis? Do you think it
would be useful to move beyond the average treatment effect (ATT) to develop a local treatment
effect (LATE)?

5. Copy your code, tables, and observations into an MS Word document and upload to the appropriate
Carmen dropbox at least 24 hours prior to our next seminar.
Instructions:
A separate set of strategic management research questions aim to describe whether or when certain events will occur. For instance, one might imagine questions such as what effects the time until the adoption of a new technology, entry into a new (non-domestic) market, or persistence of a competitive advantage. The incidence of time introduces two complications. One, tools such as logistic regression waste information (why did one firm adopt early and another late)? Two, tools such as logistic regression do not allow us to deal with censoring (how to combine data on observations that did and did not experience an event).

Event history (a.k.a., Duration or Survival) analysis refers to a collection of statistical tools that aim to describe how or explain why certain events do or do not occur. Typically, these methods deal with discrete events (e.g., market entry or survival) but can respond to sharp qualitative changes. The typical goal of the analysis is to estimate a hazard rate or (akin to) probability that an event will occur in a given period.

The objective of this problem set is to introduce you to the application of a few of the simplest event history analysis tools, the Kaplan Meir estimate and the Cox Proportional Hazard method. The assignment asks you to manipulate data from my dissertation.

Again, I do not have any preferences about the software that you use for the analysis (e.g., R, SAS, Stata) and I *encourage* you to work with others on these problem set assignments. You are simply on your honor to only associate your name with problem sets on which you have made some contribution.

1. Import the Leiblein Adoption 9096 dataset from the Carmen / Canvas course site. The New Technology Adoption dataset reports adoption of 0.05 um process technology as a function of firm, product market (e.g., Analog, ASIC, Boolean, DSP, …), and production capacity.

2. Organize the data as survival data aiming to estimate the hazard of adopting technology generation (as a function of firm characteristics).

3. How many observations survived five years? How many are censored?

4. Plot a hazard function and the survivor function assuming a Weibull and exponential distribution. Do your plots lead you to favor a parametric or partial likelihood approach to modeling this data?

5. What other functional form or modeling choices would you make for this dataset? Why?

6. Copy your code, tables, and observations into an MS Word document and upload to the appropriate Carmen dropbox at least 24 hours prior to our next seminar.