DOCTORAL SEMINAR:
CAUSAL ANALYSIS

Doctoral School
Faculty of Business and Economics
Spring 2021

Instructor:	Sirio Lonati (sirio.lonati@unil.ch)
Instructor office hours:	By appointment – online
Assistant:	Tyler Kleinbauer (tyler.kleinbauer@unil.ch)
Assistant Office hours:	By appointment – online
Class meets:	Online on Tuesday, 13h00-17h00 (but most sessions will take roughly 2 hours, from 13h00 to 15h00; Zoom meeting invitations will be sent out on a weekly basis).
Credits:	6 ECTS (outside students may audit the class with my approval)
Registration:	To register contact the Doctoral School Executive Assistant, Bénédicte Moreira (Benedicte.Moreira@unil.ch)
Website:	Moodle (password will be e-mailed to registered students)

COURSE DESCRIPTION

The use of and advanced regression methods and of structural equation modeling (SEM) have mushroomed in the past couple of decades. Today they are widely recognized as one of the most powerful and most comprehensive methods for testing causal hypotheses. Knowledge of causality is essential for informing policy and practice. The purpose of this theory and lab seminar is to familiarize the students with conceptual bases of SEM and regression, as well as applications necessary to undertake doctoral-level research and to answer questions of causal interest. Students will learn to critically think about causal relations, particularly in the design of studies, measurement of variables, and testing of theories. There will also be many demonstrations and hands-on exercises using Stata so that students have the necessary tools to analyze causal hypotheses correctly. Students will also learn basic programming commands in Stata as well as some fundamentals of Monte Carlo simulation[1].

COURSE CONTENT

In this course, students will learn about:
1. Endogeneity and causality
2. Advanced regression topics, including interpreting and testing interaction models, errors-in-variables models, instrumental variable estimators, fixed- and random-effects models, polynomial regression and response surface methodology, and simultaneous equation models,
3. SEM models including confirmatory factor analysis, latent variable models, higher-order models, multisample models, growth curve models, and model implied instrumental variables.
4. Monte Carlo analysis

---

[1] Stata is available on all office and laboratory computers. Stata 16 is also freely available for download to Unil students and employees (see https://www.unil.ch/ci/distrisoft)

The course is designed as a doctoral seminar, though a very structured one so as to maximize student learning. Students are **very strongly advised** to do assigned readings and homework, because they are essential in helping students understand the course material and being well prepared to successfully pass the project and final examination. Note, I am providing many readings from my own work or discipline, given their relevance for the course and because I can also answer detailed questions about it.

ASSESSMENT

1. 10%: homework exercises (submit homework exercises/Stata code by midday, **12h00, of the Tuesday following the relevant class to the course assistant by e-mail**); Grades for homework that are submitted late will be reduced by 1 point for each day being late (rounded to the next day).

2. 10%: Critique of articles

3. 25%: Project: Monte Carlo experiment (submitted paper and final presentation)

4. 55%: Final exam (open book exam)—theory and practice

Resit: The mark for students who fail the course will be composed of the marks received in components (1) and (2) (which cannot be redone) and a (3) resit project (i.e., this will be a new project) and/or a (4) resit final exam, weighted as per the above; note, the "and/or" depends on whether (3) or (4) or both (3) and (4) were failed. Students must resit the component/s failed; in the event that a pass cannot be obtained by resitting only (3) or (4), then both (3) and (4) must be redone.

Note on <u>formatting guidelines</u>: for the **project and article critiques**, please submit to the course assistant by e-mail using the following specifications: Double-spaced, using Times New Roman font (12 points) with default margins (2.5cm all around). Include a cover page (not counted in the page requirement) indicating your name and the type of assessment.

Important: For all work you do, please submit original work. For any project work you do or during the exam, please cite correctly and do not plagiarize; I have **failed** students in the past for plagiarism (and have a good nose for it) so please do not even let it cross your mind to use someone else's work without correct attribution. If you don't know how to cite correctly refer to a style guide like that of the APA (American Psychological Association).

PROJECT BRIEF FOR ARTICLE CRITIQUES

This project will consist of applying the concept acquired in the class to critiquing 3 papers across various management disciplines. You will be required to read the assigned papers and be prepared to discuss each of the papers. We will spend about 30-45 minutes on discussing each paper and every student will partake in the discussion.

You are also required to write max. 2 pages of critique per paper (which you must send to the course assistant by e-mail before 13h00 for each article critique session, which are on the 04-05-2021 and 25-05-2021). For each relevant week, please (a) send everything in **one** file (and not separate files), (b) follow the <u>formatting guidelines</u> for the critique (see point in previous

section), (c) clearly list (number) each of the pitfalls and each of the solutions, and (d) submit article critiques in the order in which I have listed them in the syllabus. For each paper critiqued you will be evaluated on three components:

1. Explain the basic model that was estimated by the authors (10% of mark)
2. Discuss critical errors that the authors made (40% of mark)
3. Explain how the model could be estimated to ensure correct identification of the causal effect (50% of mark).

Given the weightings the first and second component should be about a page and the last component a page also. The following paper is essential background reading for the critiques:

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. The Leadership Quarterly, 21(6). 1086-1120.

Pull out a few of the papers that Antonakis et al. (2010) coded, read them, and see what critical errors we found. Note: There may be several problems and issues with each of the papers. Please focus on identifying the major ones. <u>Do not</u> spend time talking about very minor ones like:

- they did not use a robust variance estimator
- they did not check for heteroskedasticity
- they did not rely on the chi-square test of fit in the SEM
- they used CFI or RMSEA (or the like) to evaluate model fit
- they did not correct for measurement error
- (in the case of endogenous regressors) recommend that the researchers should have used "instruments" (say rather which instruments they should have used)
- they could have used a Monte Carlo (to do what?)

Comments of the sort listed above, even if correct, will not give you <u>any</u> marks (so do not even bother making them). Look for issues that render estimates inconsistent. If you give several substantive (i.e., about 5-6) examples of problems and tangible ways to deal with them you will receive full marks for the paper critique. We will give you progressively fewer marks as the substantive content of the critique is reduced. Note, too, that describing the paper is only 20% of the mark, and the rest is on the critical errors and solutions; so please do not spend too much space on describing the basic model that was estimated. Please ensure you put the meat where the meat is needed. To get an idea of what "substantive" means, read the following paper, which on first reading may look like a very strong paper, particularly because it was published in a top journal:

Christian, M. S., & Ellis, A. P. J. (2011). Examining the effects of sleep deprivation on workplace deviance: A self-regulatory perspective. Academy of Management Journal, 54(5), 913-934.

However, all papers can be improved upon. Here are some important issues with the paper that one could have critiqued (the points below are summarized; more extended details would be required from you):

Study 1

1. The sample is self-selected, particularly to work shift; thus, there is a potential grouping variable (which is endogenous) and which is not controlled for using some sort of IV procedure.
2. The data are all self-reported data, which creates a problem of common methods variance (in addition to the untrustworthy data on the dependent variable). It would have been better to split the data-gathering and having bosses or peers report on the DVs.
3. Reverse causality is possible in the sense that those who are irritable and aggressive may sleep less; the way to deal with this is to randomize to sleep deprivation condition or to find instruments (e.g., age, personality might predict sleep).
4. The estimator is not an IV-estimator; it was estimated with OLS (and to the extent that the mediator is endogenous, failure to use an IV estimator will engender inconsistent estimates). Thus, they should have used 2SLS or ML (with disturbances correlated).

Study 2

5. There is a confound in the manipulation (i.e., sleep deprivation with putting people in a group all night long—being in a group, when in a difficult situation, could have may individuals made grumpy)—this confound induces endogeneity. Thus, the control group should have had this manipulation too (or the experimental group should not have had it).
6. The regressors are endogenous; thus, their effects on outcome must be tested with an IV estimator (as per (4) above).
7. Related to the above, the problem with testing the whole model with an IV estimator is that it is impossible to do so because the system of equations is under-identified (i.e., DF = -1); the model is thus not causally identified. Thus, they should have included more exogenous variables as instruments or manipulated a second variable (crossing it with sleep deprivation).
8. There are omitted control variables (e.g., sex, age, etc.) that are not included in the regression model. The controls are essential because of the small sample size it is possible that randomization to treatment (across sex, age, personality, what have you) is not perfect.

So, the above are examples of "meaty" comments. Thus, please ensure to make tangible and context-specific recommendations that are of substance.

PROJECT BRIEF FOR MONTE CARLO EXPERIMENT

The project consists of undertaking a Monte Carlo analysis to answer a specific question about an apparently widely accepted rule of thumb or particular practice. Examples, which can be used by students include:

- The chi-square test of fit is too powerful to assess overidentification, particularly at large sample sizes (Hu & Bentler, 1999)
- Approximate fit indices indicating RMSEA < .06 and CFI > .95 indicate a good fitting model that can be causally interpreted (Hu & Bentler, 1999)
- Factor loadings should be at least .30 for interpretable results (many references for this, but see Olatunji et al., 2007 as an example)

- Independent variables should not correlate more than .70 or .80 with each other to avoid issues related to collinearity and singularity (many textbook sources for this)
- VIFs should be less than 10 to ensure stable results (this is a widely reported rule of thumb).
- Control variables that do not correlate significantly with y should be excluded from the predictive model (cf. Becker, 2005)
- Interactions are not that much affected by common-method variance (cf. Evans, 1985)
- Common-method variance (with respect to x and y) can be eliminated by including a latent common-method factor (Loehlin, 1992).
- At least 20 observations are required for each parameter estimated in SEM models (Kline, 2010)
- A model with 10 parameters should have a sample of at least 100 (Kline, 2010)
- At least 10 observations are required for each variable in a regression model (several sources on this)
- Etc.

Note, if you are going to choose one of the examples above, please send an e-mail ASAP to the course assistant. Only one student can work on a particular topic (unless what is studied is substantially different by the two students); thus, we will operate on a first come first serve basis. You will need explicit approval from my assistant, in writing, to confirm the topic you have chosen is OK (and we will give you formal feedback on the direction you are taking with your project as indicated in the syllabus section "COURSE SCHEDULE").

The goal of the project is thus to identify some rule of thumb or suggested practice about regression or SEM, reported in a textbook, methods piece, or applied piece, and then to either debunk or to find support for it (at the least you must show something new; if you will find support for the practice or rule of thumb then do so with manipulating a dimension that has not been manipulated previously). Keep the model as simple as possible and vary sample size across a few increments and then manipulate at least two other parameters (or more) of interest. I will, of course, give you a few examples of how to do this. Note, trends may not be obvious to the naked eye; thus, you must conduct tests on the trends as a function of the manipulations (e.g., use trends as a DV and the manipulations as an IV to see what affects the DV). See the Bastardoz & Antonakis, 2016 paper for more information on how to do this—I will also show you examples in class.

The project (*and the presentation slides that you will use to present it*) must be submitted to the course assistant by e-mail before 12h00 on Tuesday 1-6-2021. Note, what you submit to me first, is final and it cannot be changed. We will mark your project on the following:

1. 10%: Introduction to the rule of thumb; literature review showing a few examples of the rule of thumb (1-2 pages max).
2. 5%: Discussion of why the rule of thumb may not or may be necessary valid for particular contexts (1 page max).
3. 25%: Set up of the Monte Carlo simulation (i.e., the program)
4. 15%: Presentation of the results of the Monte Carlo simulation (visual/graphic)
5. 20%: Discussion and interpretation of the results of the Monte Carlo simulation
6. 5%: Writing style, logic, general presentation
7. 20%: In-class oral presentation and answering of questions

The project should be between 8-10 pages long (10 pages max, excluding cover page and excluding the Appendix); please give a title to the project (e.g., "Can RSMEA be trusted to detect misspecified models?"—giving the title in question format is a good idea because it provides focus). Please include the code you used for the Monte Carlo as an Appendix in the project (not counted in the page length) and not as a separate attachment. The project should be submitted in **ONE** file.

You will present the project in class as per a randomly determined schedule. I will confirm a few weeks before how long the presentation will be (it depends on how many students will take the class, but count on a presentation of about 20 minutes or so, including Q&A).

Note about the project:
There is one simple recipe to succeed on this project—carefully follow the examples in class from the first lesson, do all the homework, and check your homework against the code we give you. If you do not understand something, ask me or the course assistant. The class gradually builds up, so ensure that you really understand everything from the first lesson. *Do not fall behind! Falling behind = high probability of failure!*

Learning basic programming skills now will help greatly for your Monte Carlo; more importantly it will help you in the long term too. Once you know how to do a Monte Carlo, this will help you in publishing empirical papers and in particular to handle reviewer comments (I will show you examples of this in class where Monte Carlo analysis helped me to publish papers that were not about Monte Carlo per se).

Finally, I or my assistant <u>will not debug</u> your code for you; the sooner you come to us for help with specific questions about your code the more we can and will help you. But please do not send us a bunch of code, screaming in desperation "my code doesn't work!" You need to do the debugging yourself because this part of the learning experience. And, finally, please do not leave the programming for the last minute; that would be a guaranteed recipe for failure. Here are some example projects to give you an idea of what has been done previously:

- Should factor loadings should be at least .30 for interpretable results?
- What is the minimum required sample size in regression models? Relative bias of two-stage least square (2SLS) and ordinary least square (OLS) estimations
- Approximate fit index RMSEA: A Monte Carlo experiment on a rule of thumb
- When the cure is worse than the disease: A Monte-Carlo experiment to test if "Common-method variance (with respect to x and y) can be eliminated by including a latent common-method factor"
- Should control variables only be included in a model under the condition that they correlate with the dependent variable?
- Is the chi-square test of fit is too powerful to assess overidentification, particularly at large sample sizes?
- Ratio of sample size to parameters for regression models
- Should independent variables not correlate more than .70 or .80 with each other to avoid issues related to collinearity and singularity?
- Impact of common-method bias on the interaction term – modelled as an omitted variable issue
- Are at least 20 observations required for each parameter estimated in SEM models?
- Should groups' ratio in a sample not exceed 1:2 allocation to avoid a serious loss of power.

- Should control variables that do not correlate significantly with y be excluded from the predictive model?
- Common-Method Variance – A test of a potentially more accurate post-hoc technique
- Does a Cronbach alpha superior (or equal) to 0.7 really absolve you from modelling measurement error?
- Can a weak instrument be detected with an F-statistic smaller than 10?
- Should control variables that do not correlate significantly with the dependent variable be excluded from the predictive model?
- Effect of sample size, degree of endogeneity, and number of indicators on the estimated coefficients in SEM
- Should VIFs be less than 10?
- Factor loadings: investigating the rules of thumb
- Is it ever justified to use a Dawes-model for predictions?
- Weak instruments: Is more better?
- Detecting Weak Instruments: Avoiding a cure worse than the disease
- Fast and Frugal Regression: Take-The-Best competing with multiple regression
- When do control variables matter?
- What constitutes a sufficient sample size for accurate estimation in multilevel modeling?
- Endogeneity tests: Which one to trust?
- Can model implied instruments reliably estimate latent interactions and latent quadratic terms?
- The marker variable technique: A relevant statistical technique to control for CMV?

COURSE DELIVERY

Given the limitations imposed by the COVID pandemic, **the entire seminar will be held remotely**; due to the exceptional circumstances, however, changes to the class structure and/or to its delivery are always possible. Throughout the semester, we will have two types of activities (see Figure 1 for an overview of the class' structure).

New topics every week: Pre-recorded videos, slides, and readings. In each week, I will introduce a new topic and I will post videos and other class materials (e.g., readings, slides). New materials will be generally available on Tuesdays and will be accessible remotely through the Moodle and other Unil's platforms. These materials can be accessed at any time and at your own individual speed.

Interactive sessions. Each week, students will also have the chance to ask clarification questions in two different occasions:

1. <u>Real-time sessions.</u> We will have real-time sessions on Zoom (which is available for download on https://www.unil.ch/ci/distrisoft), where we will briefly *recap the material posted in the preceding week* and where students will have the *chance to ask questions to the instructor and to the guest lecturer*. To get the most out of these real-time sessions, you are required to familiarize yourself with the materials posted prior to each real-time session. In some of these sessions, students will also present their Monte Carlo projects and their article critiques. With a few exceptions, most sessions will take approximately 2 hours, depending on the weekly material, on the size of the class, and on the number of questions you will have.

2. <u>A Moodle forum will always be available.</u> Students can ask questions directly on the Moodle forum, where other students or the course assistant will reply. The idea is to use this public forum to share knowledge and insights, so that each of you can benefit also from your colleagues' questions. *Note: On a general basis, the course instructor and the course assistant will not answer clarification questions on the class' material sent via email.*
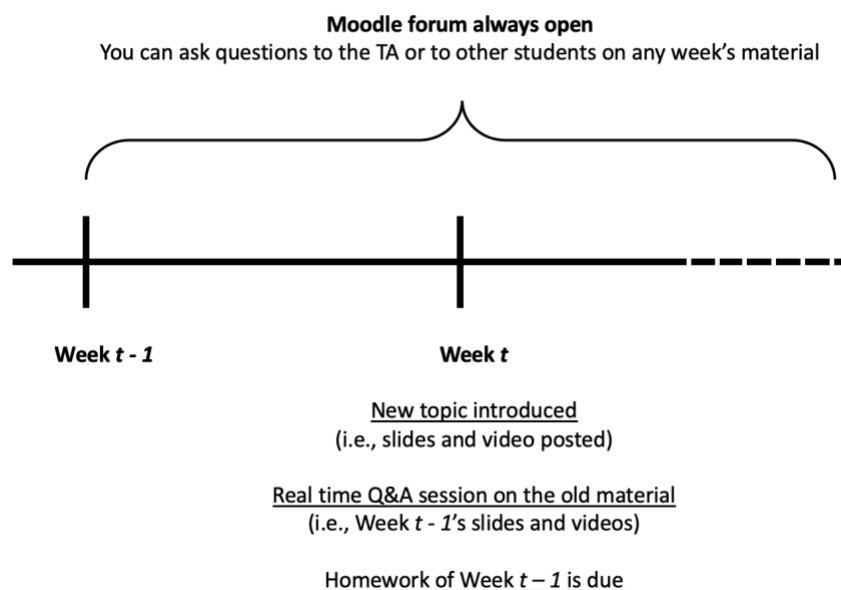

Figure 1. General structure of the seminar on a weekly basis

COURSE SCHEDULE

Note: because of the current pandemic, to maximize student learning outcomes I reserve the right to change schedule, readings, and lab topics at any time during the course.

<u>*Monday 22-2-2021*</u> (optional lesson) — <u>13h00-17h00 (live on Zoom: I will send you the Zoom link some days before the session)</u>

Note:	You should come to this class if you have never used Stata and particularly if you have not had any courses previously in econometrics. Students who have studied in Lausanne for their BSc and MSc and/or have taken many econometrics classes do not need to attend this lesson, but of course are welcome to if they wish (it may serve as a good recap; thus, it would be a good idea for all students to attend in any case). To get the most of this session, you should have a fully functional version of Stata on your laptop.
New topic:	Introduction to testing causal hypotheses; experimental research; introduction to endogeneity; introduction to SEM and regression. Basic notations for regression and path diagrams; introduction to Stata.
Required reading:	1. Jacquart, P., Cole, M. S., Gabriel, A. S., Koopman, J., & Rosen, C. C. 2017. Studying leadership: Research design and methods. In J. Antonakis & D. V. Day (Eds.), The Nature of Leadership, 3 ed.: 411-437. Thousand Oaks: Sage. (this chapter is written in a general way and understandable to individuals from various disciplines).
Recommended:	1. Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. Journal of Operations Management, 64, 19-40. (for those of you interested in experimental research)
Lab topics:	Familiarization with Stata; basic data manipulation; estimating basic models.

Review the following commands by typing "help [command]"; then run the commands using the systems data set that is noted in the help file. Note, many of the commands can be abbreviated; and many of the commands can be used with the menus. We will use the following commands:

Basic operations with respect files: clear, use, edit, sysuse, save, preserve, restore
Basic operations with respect to variables: describe, list, summarize, codebook, generate, label, replace, egen, keep, drop, recode, tabulate, tabulate (for generating dummy variables, i.e., "tab country, gen(d)")
Basic operations to manage data: sort, gsort, order, reshape, merge
Basic operations with respect to estimation: pwcorr, corr, ttest, anova, regress, logit, probit
Keeping track of your work: do files, log files.
Installing user-written commands: ssc install, findit

Live session:	We will have our first live session from 13h00 to 15h00 (the Zoom invitation link will be sent via the Moodle). I will introduce the course, explain the grading scheme, and present the class structure. I will also post the videos and materials related to the first week (see the section "New topic" below).
New topic:	Introduction to the course. The problem of endogeneity: 1. Basic omitted variable bias 2. Errors-in-variables (using least squares and SEM) 3. Common-method variance
Required reading:	1. Jacquart, P., Cole, M. S., Gabriel, A. S., Koopman, J., & Rosen, C. C. 2017. Studying leadership: Research design and methods. In J. Antonakis & D. V. Day (Eds.), The Nature of Leadership, 3 ed.: 411-437. Thousand Oaks: Sage. (this chapter is written in a general way and understandable to individuals from various disciplines). 2. Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. In D.V. Day (Ed.), The Oxford Handbook of Leadership and Organizations. Read pages 1-29 only. 3. Schwarz, N. 1999. Self-reports - How the questions shape the answers. American Psychologist, 54(2): 93-105.
Recommended:	1. Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. The Leadership Quarterly, 21(6). 1086-1120. Read pages 1086-1092 (up to section 3.1.1); Section 3.3, 3.4,
Lab topics:	Generating data with known structures (generate a data set in which x is endogenous). Generating data with loops. We will use the following commands or command options (-[option]-): set obs, set seed, gen, -rnormal()-, foreach, regress, alpha, eivreg, sem, est store, est tab, esttab, outreg2, putdocx

Week 2; Tuesday 2-3-2021

Live session:	We will have a live session held on Zoom to clarify Week 1's material (i.e., "The problem of endogeneity") at 13h00-15h00.
New topic:	Regression models in Stata 1. Multiple IVs and dummy variables (and coefficients); relation to ANOVA 2. Nested model F-test and hypothesis tests (Wald tests) 3. Marginal effects 4. Interactions (estimating and plotting)

5. Testing congruence (brief introduction)—see notes on Moodle
6. Regression diagnostics

Required reading:
(Practical example)   1. Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. Journal of Business and Psychology, 29(1), 1-19.

Suggested readings for congruence testing (for those interested):
(Practical example)   1. Lee, Y. T., & Antonakis, J. (2014). When Preference Is Not Satisfied but the Individual Is: How Power Distance Affects Person-Job Fit. Journal of Management, 40(3), 641-675. (Scan this to see what can easily be done with Stata with respect to testing pretty much anything; see supplementary notes on the Moodle for more information if you need to test response surfaces).

2. Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores. Journal of Business and Psychology, 25(4), 543-554.

Lab topics: Estimating regression models in practice. We will use the following commands or command options (-[option]-):

oneway, anova, regress, hausman, suest, lincom, test, bootstrap, margins, -robust-,

Week 3; Tuesday 9-3-2021

Live session: We will have a live session held on Zoom on Week 2's material (i.e., "Regression models in Stata") at 13h00-15h00.

New topic: Two-stage equation models in Stata:
1. Two-stage least squares (instrumental variable) estimator
2. Tests for overidentifying restrictions; what the chi-square test means
3. Hausman test for endogeneity and augmented regression tests (Durbin–Wu–Hausman)
4. Tests of mediation (Sobel-Goodman Tests)
5. Bootstrapping

Required reading: 1. Sajons, G. (2019). Estimating the causal effect of measured endogenous variables: A tutorial on the experimental instrumental variable approach. _The Leadership Quarterly_.

2. Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. _The Leadership Quarterly_, 21(6). 1086-1120. Read 4.2.1, 4.2.1.4 (Example 1 only)

3. Lonati, S. (2020). What explains cultural differences in leadership styles? On the agricultural origins of participative and directive leadership. *The Leadership Quarterly*,31(2), 101305

Suggested reading: Refer to one or two of the following, which is closest to your field:

1. Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, 6(3), 285-327.
2. Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, 49(3), 186-205.
3. Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, 1(1), 51-78.
4. Duncan, G. J., Magnusson, K. A., & Ludwig, J. (2004). The Endogeneity Problem in Developmental Studies. *Research in Human Development*, 1(1&2), 59-80.
5. Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology*, 44(2), 381-394.
7. Bettis, R., Gambardella, A., Helfat, C., & Mitchell, W. (2014). Editorial: Quantitative empirical analysis in strategic management. *Strategic Management Journal*, 35(7), 949-953.
8. Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2020). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, 0149206320960533.
9. Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, 52, 1-14.
10. Maula, M., & Stam, W. (2019). Enhancing Rigor in Quantitative Entrepreneurship Research. *Entrepreneurship Theory and Practice*, https://doi.org/10.1177/1042258719891388
11. Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, 25(2), 243-258.
12. Reeb, D., Sakakibara, M., & Mahmood, I. P. (2012). From the Editors: Endogeneity in international business research. *Journal of International Business Studies*, 43(3), 211-218.
13. Shaver, J. M. (2019). Causal Identification Through a Cumulative Body of Research in the Study of Strategy and Organizations. *Journal of Management*, DOI: 10.1177/0149206319846272.
14. Ullah, S., Zaefarian, G., & Ullah, F. (in press). How to use instrumental variables in addressing endogeneity? A step-by-step procedure for non-specialists. DOI: https://doi.org/10.1016/j.indmarman.2020.03.006

15. Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2021). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, 47(1), 105-143.

Lab topics: We will estimate models using the following commands: ivregress, ivreg2, reg3, sem, sgmediation, test, nlcom, hausman, bootstrap

Week 4; Tuesday 16-3-2021

Live session: We will have a live session held on Zoom on Week 3's material (i.e., "Two-stage equation models in Stata") at 13h00-15h00. (*Instructor: Sirio Lonati; Guest lecturer: John Antonakis*)

New topic: 1. Simultaneous equation models and seemingly unrelated models
- a. Multiple dependent variables
- b. Seemingly-unrelated regression/estimation
- c. Path models with three-stage least squares
- d. Path models with sem
- e. Simultaneity
- f. Cross-equation Wald tests and Chow tests

2. Path and SEM models:
- a. ML estimation and assumptions
- b. SEM discrepancy function
- c. Calculating the chi-square manually for path (IV) models
- d. Likelihood ratio tests (for nested models)

Required reading: 1. Bollen, K. A. (1998). Structural equation models. John Wiley & Sons, Ltd. Read Chapter 2, pp. 10-39.

2. MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. Annual Review of Psychology, 51(1), 201-226.

Lab topics: We will estimate model using the following commands: sureg, cmp, reg3, sem, suest, test

Week 5; Tuesday 23-3-2021

Live session: We will have a live session held on Zoom on Week 4's material (i.e., "Simultaneous equation, SUR, path and SEM models") at 13h00-15h00. (*Instructor: Sirio Lonati*)

New topic: Confirmatory factor analysis and advanced topics
1. CFA models, basic and multifactor; calculating chi-square
2. Evaluating fit: fit statistics, residuals, modification indices (Lagrange tests)
3. SEM model with latent variables
4. SEM Mediation models

	5. Likelihood ratio tests and Wald tests for parameter constraints
	6. Higher-order factor models

Required reading:	1. Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21(6). 1086-1120. Read 4.2.1.4 (Example 2 only)
	Tomarken, A. J., & Waller, N. G. (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions. Annual Review of Clinical Psychology, 1(1), 31-65.
Suggested reading: (Practical examples)	1. Lee, Y. T., Stettler, A., & Antonakis, J. (2011). Incremental Validity and Indirect effect of Ethical Development on Work Performance. *Personality and Individual Differences*, 50(7), 1110-1115.
	2. Fiori, M., & Antonakis, J. (2011). The ability model of emotional intelligence: Searching for valid measures. *Personality and Individual Differences*, 50(3), 329-334.
Lab topics:	We will estimate models using the following commands: sem (and various sem options). We will also use some of the basic matrix commands of Stata (determinants, trace, inverse)

Week 6; Tuesday 30-3-2021

Live session:	We will have a live session held on Zoom on Week 5's material (i.e., "Confirmatory factor analysis and advanced topics") at 13h00-15h00.
New topic:	Advanced path and SEM models in Stata with latent variables I 1. A note on fit and a bit more on Monte Carlo 2. On local and global fit 3. Model implied instrumental variables with ML and 2sls 4. Latent variable interaction models
Required reading :	1. Ropovik, I. 2015. A cautionary note on testing latent variable models. *Frontiers in Psychology*, 6(1715).
Suggested reading: (for those interested)	1. Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least-squares estimator. *Sociological methodology*, 25, 223-252.
	2. Bollen, K. A., & Paxton, P. (1998). Interactions of latent variables in structural equation models. Structural Equation Modeling: A Multidisciplinary Journal, 5(3), 267-293.
	Lab topics: We will estimate model using the following commands: sem (and various sem options), ssd.

Live session: We will have a live session held on Zoom on Week 6's material (i.e., "Advanced path and SEM models in Stata with latent variables I") at 13h00-15h00.

New topic: Advanced path and SEM models in Stata with latent variables II
1. Multisample (multiple group) models (same and different DVs) and tests of invariance
2. MIMIC models
3. Seemingly unrelated SE models (same and different DVs)
4. Analyzing secondary data (using covariance) matrixes (with ssd and corr2data)

Required reading: 1. Antonakis, J. (2009). "Emotional intelligence": What does it measure and does it matter for leadership? In G. B. Graen (Ed.), *LMX leadership--Game-Changing Designs: Research-Based Tools* (Vol. Vol. VII, pp. 163-192). Greenwich, CT: Information Age Publishing.

Suggested reading: 1. Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 3*(1), 4-70.

Week 8; Tuesday 20-4-2021

Live session: We will have a live session held on Zoom on Week 7's material (i.e., "Advanced path and SEM models in Stata with latent variables II") at 13h00-15h00.

New topic: Panel data, SEM style
1. Fixed-effects models
2. Random-effects model
3. Combining fixed- and random-effects (the "Mundlak" estimator)
4. Growth curve models

Required reading: 1. Antonakis, J., Bastardoz, N., & Rönkkö, M. (2019). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, 1094428119877457.

2. Bollen, K. A., & Brand, J. E. (2010). A General Panel Model with Random and Fixed Effects: A Structural Equations Approach. Social Forces, 89(1), 1-34.

Suggested reading: 1. Bou, J. C., & Satorra, A. (2018). Univariate Versus Multivariate Modeling of Panel Data: Model Specification and Goodness-of-Fit Testing. *Organizational Research Methods*, 21(1), 150-196.

2. McNeish, D., Stapleton, L. M., & Silverman, R. D. 2016. On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*: http://dx.doi.org/10.1037/met0000078.

3. Bliese, P. D., Schepker, D. J., Essman, S. M., & Ployhart, R. E. (2020).
Bridging Methodological Divides Between Macro- and Microresearch: Endogeneity and Methods for Panel Data. *Journal of Management*, 46(1), 70-99.

4. McNeish, D., & Kelley, K. (2018). Fixed effects models versus mixed
effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24(1), 20-35.

Practical example:	1. Bendahan, S., Zehnder, C., Pralong, F. P., & Antonakis, J. 2015. Leader corruption depends on power and testosterone. *The Leadership Quarterly, 26,* 101-122. (See Study 2 regarding how a SEM growth curve analysis was used in the context of an experiment)
Lab topics:	We will estimate models with Stata using xtreg, and regress and also use the commands lincom and margins -cluster(id)-, _n, expand, and sem

Week 9; Tuesday 27-4-2021

Live session:	We will have a live session held on Zoom on Week 8's material (i.e., "Panel data") at 13h00-15h00.
New topic:	Monte Carlo simulations.
Required reading:	1. Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. N. (2001). Monte Carlo Experiments: Design and Implementation. Structural Equation Modeling-a Multidisciplinary Journal, 8(2), 287-312.

2. Muthen, L. K., & Muthen, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 9(4), 599-620. |
| Practical example: | 1. Bastardoz, N. & Antonakis, J. 2016. How should the fit of structural equation models be judged? Insights from Monte Carlo simulations. Academy of Management Proceedings, 12634. |
| Optional reading: (for examples) | 1. Mewhort, D. J. K. (2005). A comparison of the randomization test with the F test when error is skewed. Behavior Research Methods, 37(3), 426-435.

2. Semadeni, M., Withers, M. C., & Certo, S. T. 2014. The perils of |

endogeneity and instrumental variables in strategy research: Understanding through simulations. Strategic Management Journal, 35(7): 1070-1079.

Lab topics: We will review code to generating data structures, including drawnorm, and basic code to write programs

Week 10; Tuesday 4-5-2021

Live session: We will have a live session held on Zoom on Week 9's material (i.e., "Monte Carlo simulations") at 13h00-15h00. From 15h00 to 16h00/17h00 we will have our first article critique.

New topic: Critique of 1 article

Articles to critique: 1. Erez, A., & Judge, T. A. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology*, 86(6), 1270.

Suggested reading: 1. Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. The Leadership Quarterly, 28(1), 5-21.

Week 11; Tuesday 11-5-2021

Live session: We will have a live session held on Zoom from 13h00 to 17h00. Please note that this class might take a bit more than 4 hours. (*Instructor: Sirio Lonati; Guest lecturer: John Antonakis*)

You need to prepare a very brief overview of your Monte Carlo experiment. Please e-mail it to Tyler before the start of the lesson; you are required to have only 3-4 slides on the Monte Carlo, which you will present in front of the class. On the first slide discuss the rule of thumb and whether you think it is baseless or a sound advice; on the second slide, discuss which variables you will manipulate for the Monte Carlo and why, and the basic setup of the Monte Carlo. On the third slide, present the basic code to show what you will manipulate (i.e., this code is to show only the data generation—highlighting the parameters you will manipulate)

New topic: No new topic will be introduced.

Week 12; Tuesday 18-5-2021 (Instructor: Sirio Lonati; Guest lecturer)

Sirio, and Tyler will meet with each of you, privately, during the week at a mutually convenient time to discuss the Monte Carlo project. We will set up a meeting schedule with you on Doodle, which we will communicate to you a couple of weeks before the meeting. If we need more time to meet all students, meetings may go beyond the allotted class time (i.e., we may plan additional sessions on Tuesday or on other days).

New topic: Critique of 2 articles

Articles to critique: 1. Kakkar, H., & Sivanathan, N. (2017). When the appeal of a dominant leader is greater than a prestige leader. *Proceedings of the National Academy of Sciences*, 114(26), 6734-6739.

2. Deng, H., Wu, C. H., Leung, K., & Guan, Y. (2016). Depletion from self-regulation: A resource-based account of the effect of value incongruence. *Personnel Psychology*, 69(2), 431-465.

*this lesson will be held live on Zoom, and will take around 2 hours long, depending on amount of discussion we have. I will be available for final (pointed) questions regarding your Monte Carlo presentations if required. Else you can use the time to work on your Monte Carlos.

Week 14; Tuesday 1-6-2021, and possibly Wednesday 2-6-2021 (Instructor: Sirio Lonati; Guest lecturer: John Antonakis)

Monte Carlo project presentations held live on Zoom. Note, this session will likely start early and run late, depending on how many students are required to present. If we do not have sufficient slots in which to accommodate all students on this day, we may use Wednesday, too.

Week 15: To be determined (2021).
Final exam (most likely virtual, to be determined)

References:

Becker, T. E. (2005). Potential Problems in the Statistical Control of Variables in Organizational Research: A Qualitative Analysis With Recommendations. *Organizational Research Methods, 8*(3), 274-289.

Evans, M. G. (1985). A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis. *Organizational Behavior and Human Decision Processes, 36*, 305-323.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., et al. (2007). The disgust scale: Item analysis, factor structure, and suggestions for refinement. *Psychological Assessment, 19*(3), 281-297.

SUGGESTED REFERENCE BOOKS FOR YOUR PERSONAL LIBRARIES
(PLEASE SERIOUSLY CONSIDER BUYING SOME OF THESE BOOKS)

Stata and regression-related books:
Acock, A. C. (2016). A gentle introduction to Stata (5th edition). College Station, TX: Stata Press--gives a nice overview to Stata for those who have never used it (for basic use).

Baum, C. F. (2006). An introduction to modern econometrics using Stata. College Station, TX.: Stata Press.--a very useful general purpose reference guide for using Stata (for intermediatry-advanced use)

Cameron, A. C., & Trivedi, P. K. (2009). Microeconometrics Using Stata. College Station, TX: Stata Press. --an extremely useful general purpose reference guide for using Stata (for advanced use)

Long, J. S., & Freese, J. (2006). Regression models for categorical dependent variables using Stata (2nd ed.). College Station, TX: StataCorp LP. --an extremely useful reference guide for categorical dependent variables using Stata (for advanced use)

SEM and Mplus-related books:
Byrne, B. M. (2012). Structural equation modelling with Mplus. New York: Routledge.

Muthén, L. K., & Muthén, B. O. (2010). Mplus user's guide (6th ed.). Los Angeles, CA: Muthén & Muthén. http://www.statmodel.com/ugexcerpts.shtml

Muthén, B., Muthén, L. K., & Asparouhov, T. 2016. Regression and mediation analysis using Mplus: Los Angeles, CA: Muthén & Muthén.

General books on SEM
Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley & Sons.--this book is **essential**, a classic, and will give you lots of information regarding SEM in general. It is quite advanced, so refer to this only for more detailed information on SEM issues. And, do not be fooled by the fact that it is dated; the book is timeless.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford Press.--basic introduction and reference guide.

Duncan, T. E., Duncan, S. C., Stryker, L. A., Li, F., & Alpert, A. (1999). An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah, NJ: Lawrence Erlbaum Associates.--as denoted in the title, this book is essential reading for growth-curve models.

Heck, R.H., & Thomas, S.L. (2000). An introduction of multilevel modeling techniques. Mahwah, NJ: Lawrence Erlbaum Associates. --as denoted in the title, this book is essential reading for multilevel models.

Kline, R. B. 2015. Principles and practice of structural equation modeling (4th ed.). New York: Guilford Press.

Loehlin, J. C. (1992). Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, NJ: Lawrence Erlbaum.--basic introduction and reference guide.

Maruyama, G. M. (1998). Basics of structural equation modeling. Thousand Oaks, CA: Sage Publications.--basic introduction and reference guide.

Note: The following site has useful information on Mplus and Stata:
http://www.ats.ucla.edu/stat/seminars/

As regards Stata, the following are useful:
http://www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm
http://www.stata.com/support/faqs/

Stata-related books can be obtained on-line from www.stata.com or locally from http://www.scientific-solutions.ch/.

Note: for those of you who need to brush up on regression, see:

Angrist, J. D. & Pischke, J.-S. 2008. Mostly harmless econometrics: An empiricist's companion. Princeton: Princeton University Press.

Angrist, J. D. & Pischke, J.-S. 2014. Mastering 'metrics: The path from cause to effect. Princeton: Princeton University Press.

Stock, J. H., & Watson, M. W. (2007). Introduction to econometrics (2nd ed.). Boston: Pearson Addison Wesley.

Wooldridge, J. M. 2013. Introductory econometrics: A modern approach (5th ed.). Mason, OH: South-Western Cengage Learning.

Note: for those of you who need to brush up on experimental design, see:

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.