

CONTENTS



List of Figures vii

List of Tables ix

Preface xi

Acknowledgments xv

Organization of This Book xvii

I PRELIMINARIES 1

1 Questions about *Questions* 3

2 The Experimental Ideal 11

2.1 The Selection Problem 12

2.2 Random Assignment Solves the Selection Problem 15

2.3 Regression Analysis of Experiments 22

II THE CORE 25

3 Making Regression Make Sense 27

3.1 Regression Fundamentals 28

3.2 Regression and Causality 51

3.3 Heterogeneity and Nonlinearity 68

3.4 Regression Details 91

3.5 Appendix: Derivation of the Average Derivative
Weighting Function 110

4 Instrumental Variables in Action: Sometimes
You Get What You Need 113

4.1 IV and Causality 115

4.2 Asymptotic 2SLS Inference 138

4.3 Two-Sample IV and Split-Sample IV 147

vi Contents

4.4	IV with Heterogeneous Potential Outcomes	150
4.5	Generalizing LATE	173
4.6	IV Details	188
4.7	Appendix	216
5	Parallel Worlds: Fixed Effects, Differences-in-Differences, and Panel Data	221
5.1	Individual Fixed Effects	221
5.2	Differences-in-Differences	227
5.3	Fixed Effects versus Lagged Dependent Variables	243
5.4	Appendix: More on Fixed Effects and Lagged Dependent Variables	246
III	EXTENSIONS	249
6	Getting a Little Jumpy: Regression Discontinuity Designs	251
6.1	Sharp RD	251
6.2	Fuzzy RD Is IV	259
7	Quantile Regression	269
7.1	The Quantile Regression Model	270
7.2	IV Estimation of Quantile Treatment Effects	283
8	Nonstandard Standard Error Issues	293
8.1	The Bias of Robust Standard Error Estimates	294
8.2	Clustering and Serial Correlation in Panels	308
8.3	Appendix: Derivation of the Simple Moulton Factor	323
	<i>Last Words</i>	327
	<i>Acronyms and Abbreviations</i>	329
	<i>Empirical Studies Index</i>	335
	<i>References</i>	339
	<i>Index</i>	361

Chapter 1

Questions about *Questions*



“I checked it very thoroughly,” said the computer, “and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you’ve never actually known what the question is.”

Douglas Adams, *The Hitchhiker’s Guide to the Galaxy*

This chapter briefly discusses the basis for a successful research project. Like the biblical story of Exodus, a research agenda can be organized around four questions. We call these frequently asked questions (FAQs), because they should be. The FAQs ask about the relationship of interest, the ideal experiment, the identification strategy, and the mode of inference.

In the beginning, we should ask, *What is the causal relationship of interest?* Although purely descriptive research has an important role to play, we believe that the most interesting research in social science is about questions of cause and effect, such as the effect of class size on children’s test scores, discussed in chapters 2 and 6. A causal relationship is useful for making predictions about the consequences of changing circumstances or policies; it tells us what would happen in alternative (or “counterfactual”) worlds. For example, as part of a research agenda investigating human productive capacity—what labor economists call human capital—we have both investigated the causal effect of schooling on wages (Card, 1999, surveys research in this area). The causal effect of schooling on wages is the increment to wages an individual would receive if he or she got more schooling. A range of studies suggest the causal effect of a college degree is about 40 percent higher wages on average, quite a payoff. The causal

effect of schooling on wages is useful for predicting the earnings consequences of, say, changing the costs of attending college, or strengthening compulsory attendance laws. This relation is also of theoretical interest since it can be derived from an economic model.

As labor economists, we're most likely to study causal effects in samples of workers, but the unit of observation in causal research need not be an individual human being. Causal questions can be asked about firms or, for that matter, countries. Take, for example, Acemoglu, Johnson, and Robinson's (2001) research on the effect of colonial institutions on economic growth. This study is concerned with whether countries that inherited more democratic institutions from their colonial rulers later enjoyed higher economic growth as a consequence. The answer to this question has implications for our understanding of history and for the consequences of contemporary development policy. Today, we might wonder whether newly forming democratic institutions are important for economic development in Iraq and Afghanistan. The case for democracy is far from clear-cut; at the moment, China is enjoying robust economic growth without the benefit of complete political freedom, while much of Latin America has democratized without a big growth payoff.

The second research FAQ is concerned with *the experiment that could ideally be used to capture the causal effect of interest*. In the case of schooling and wages, for example, we can imagine offering potential dropouts a reward for finishing school, and then studying the consequences. In fact, Angrist and Lavy (2008) have run just such an experiment. Although their study looked at short-term effects such as college enrollment, a longer-term follow-up might well look at wages. In the case of political institutions, we might like to go back in time and randomly assign different government structures in former colonies on their independence day (an experiment that is more likely to be made into a movie than to get funded by the National Science Foundation).

Ideal experiments are most often hypothetical. Still, hypothetical experiments are worth contemplating because they help us pick fruitful research topics. We'll support this claim by

asking you to picture yourself as a researcher with no budget constraint and no Human Subjects Committee policing your inquiry for social correctness: something like a well-funded Stanley Milgram, the psychologist who did pathbreaking work on the response to authority in the 1960s using highly controversial experimental designs that would likely cost him his job today.

Seeking to understand the response to authority, Milgram (1963) showed he could convince experimental subjects to administer painful electric shocks to pitifully protesting victims (the shocks were fake and the victims were actors). This turned out to be controversial as well as clever: some psychologists claimed that the subjects who administered shocks were psychologically harmed by the experiment. Still, Milgram's study illustrates the point that there are many experiments we can think about, even if some are better left on the drawing board.¹ If you can't devise an experiment that answers your question in a world where anything goes, then the odds of generating useful results with a modest budget and nonexperimental survey data seem pretty slim. The description of an ideal experiment also helps you formulate causal questions precisely. The mechanics of an ideal experiment highlight the forces you'd like to manipulate and the factors you'd like to hold constant.

Research questions that cannot be answered by any experiment are FUQs: fundamentally unidentified questions. What exactly does a FUQ look like? At first blush, questions about the causal effect of race or gender seem good candidates because these things are hard to manipulate in isolation ("imagine your chromosomes were switched at birth"). On the other hand, the issue economists care most about in the realm of race and sex, labor market discrimination, turns on whether someone treats you differently because they *believe* you to be black or white, male or female. The notion of a counterfactual world where men are perceived as women or vice versa has a long history and does not require Douglas Adams-style outlandishness to entertain (Rosalind disguised

¹Milgram was later played by the actor William Shatner in a TV special, an honor that no economist has yet received, though Angrist is still hopeful.

as Ganymede fools everyone in Shakespeare's *As You Like It*). The idea of changing race is similarly near-fetched: in *The Human Stain*, Philip Roth imagines the world of Coleman Silk, a black literature professor who passes as white in professional life. Labor economists imagine this sort of thing all the time. Sometimes we even construct such scenarios for the advancement of science, as in audit studies involving fake job applicants and résumés.²

A little imagination goes a long way when it comes to research design, but imagination cannot solve every problem. Suppose that we are interested in whether children do better in school by virtue of having started school a little older. Maybe the 7-year-old brain is better prepared for learning than the 6-year-old brain. This question has a policy angle coming from the fact that, in an effort to boost test scores, some school districts are now imposing older start ages (Deming and Dynarski, 2008). To assess the effects of delayed school entry on learning, we could randomly select some kids to start first grade at age 7, while others start at age 6, as is still typical. We are interested in whether those held back learn more in school, as evidenced by their elementary school test scores. To be concrete, let's look at test scores in first grade.

The problem with this question—the effects of start age on first grade test scores—is that the group that started school at age 7 is . . . older. And older kids tend to do better on tests, a pure maturation effect. Now, it might seem we can fix this by holding age constant instead of grade. Suppose we wait to test those who started at age 6 until second grade and test those who started at age 7 in first grade, so that everybody is tested at age 7. But the first group has spent more time in school, a fact that raises achievement if school is worth anything. There is no way to disentangle the effect of start age on learning from maturation and time-in-school effects as long as kids are still in school. The problem here is that for students, start age

²A recent example is Bertrand and Mullainathan (2004), who compared employers' responses to résumés with blacker-sounding and whiter-sounding first names, such as Lakisha and Emily (though Fryer and Levitt, 2004, note that names may carry information about socioeconomic status as well as race.)

equals current age minus time in school. This deterministic link disappears in a sample of adults, so we can investigate pure start-age effects on adult outcomes, such as earnings or highest grade completed (as in Black, Devereux, and Salvanes, 2008). But the effect of start age on elementary school test scores is impossible to interpret even in a randomized trial, and therefore, in a word, FUQed.

The third and fourth research FAQs are concerned with the nuts-and-bolts elements that produce a specific study. Question number 3 asks, *What is your identification strategy?* Angrist and Krueger (1999) used the term *identification strategy* to describe the manner in which a researcher uses observational data (i.e., data not generated by a randomized trial) to approximate a real experiment. Returning to the schooling example, Angrist and Krueger (1991) used the interaction between compulsory attendance laws in American states and students' season of birth as a natural experiment to estimate the causal effects of finishing high school on wages (season of birth affects the degree to which high school students are constrained by laws allowing them to drop out after their 16th birthday). Chapters 3–6 are primarily concerned with conceptual frameworks for identification strategies.

Although a focus on credible identification strategies is emblematic of modern empirical work, the juxtaposition of ideal and natural experiments has a long history in econometrics. Here is our econometrics forefather, Trygve Haavelmo (1944, p. 14), appealing for more explicit discussion of both kinds of experimental designs:

A design of experiments (a prescription of what the physicists call a “crucial experiment”) is an essential appendix to any quantitative theory. And we usually have some such experiment in mind when we construct the theories, although—unfortunately—most economists do not describe their design of experiments explicitly. If they did, they would see that the experiments they have in mind may be grouped into two different classes, namely, (1) experiments that *we should like to make* to see if certain real economic phenomena—when artificially isolated from “other influences”—would verify certain

hypotheses, and (2) the stream of experiments that Nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers. In both cases the aim of the theory is the same, to become master of the happenings of real life.

The fourth research FAQ borrows language from Rubin (1991): *What is your mode of statistical inference?* The answer to this question describes the population to be studied, the sample to be used, and the assumptions made when constructing standard errors. Sometimes inference is straightforward, as when you use census microdata samples to study the American population. Often inference is more complex, however, especially with data that are clustered or grouped. The last chapter covers practical problems that arise once you've answered question number 4. Although inference issues are rarely very exciting, and often quite technical, the ultimate success of even a well-conceived and conceptually exciting project turns on the details of statistical inference. This sometimes dispiriting fact inspired the following econometrics haiku, penned by Keisuke Hirano after completing his thesis:

*T-stat looks too good
Try clustered standard errors—
Significance gone*

As should be clear from the above discussion, the four research FAQs are part of a process of project development. The following chapters are concerned mostly with the econometric questions that come up after you've answered the research FAQs—in other words, issues that arise once your research agenda has been set. Before turning to the nuts and bolts of empirical work, however, we begin with a more detailed explanation of why randomized trials give us our benchmark.

INDEX

- Abadie, Alberto, 77n, 171n, 178, 181n, 198, 244n, 284, 286
- Abadie kappa (weighting function), 172n, 179–81, 203–5, 286–7
- ability bias and returns to schooling, 60–62, 66–68
- absorbing fixed effects, 224
- Acemoglu, Daron, 4, 28, 166, 184–85, 193–94
- Achilles, Charles M., 17
- Ackerberg, Daniel A., 210n
- Adams, Douglas, 3, 11, 27, 113, 221, 251, 269, 293, 319; *Life, the Universe, and Everything*, 316n
- AFQT. *See* Armed Forces Qualification Test
- Altonji, Joseph G., 93n
- always-takers, 158–61
- Amemiya, Takeshi, 142n
- Ammermueller, Andreas, 196
- analysis of covariance, 224
- analysis of variance (ANOVA), 33–34
- Ananat, Elizabeth, 137
- Anderson, Michael, 12n
- Angrist, Joshua D., 4, 5n, 7, 16, 21, 22n, 23, 27–28, 40, 63, 69, 71–72, 78–79, 84–85, 94, 101, 105, 115n, 117–20, 123, 126–132, 136–38, 146n, 148–50, 148n, 149n, 153–56, 153n, 156n, 160, 166–70, 172, 174, 175, 177, 180, 182–84, 186, 193, 196, 202, 205n37, 207n, 210n, 213–14, 215n, 241, 256–57, 263–64, 275n3, 278–79, 284, 286–87, 308, 311, 314n, 321, 322
- ANOVA theorem, 33–34
- Arellano, Manuel, 245n, 319n
- Armed Forces Qualification Test (AFQT), 61, 71, 189–90
- Ashenfelter, Orley, 16, 21, 86, 226–27, 243–44
- asymptotic covariance matrix, OLS, 44–48, 294–300; 2SLS, 138–40
- asymptotic refinement, 301–2
- ATE. *See* treatment effects: unconditional average (ATE)
- Athey, Susan, 230n
- Atkinson, Anthony B., 285
- Austin, Gerald, 305n
- autocorrelation. *See* serial correlation
- Autor, David, 34, 237–38, 274
- average causal effect. *See* treatment effects
- average causal response, 182–88
- average derivative, 77–80, 104, 110, 187–88
- average treatment effect. *See* treatment effects
- bad control, 64–68
- Baker, Regina, 120n, 205n39, 213
- Barnett, Steven W., 11
- Barnow, Burt S., 53, 59
- Bassett, Gilbert, 272, 284
- Becker, Sascha, 83n

- Behrens-Fisher problem, 302–8
Bekker, Paul A., 205n39, 207n, 210n
Bell, Robert M., 320–21, 320n
Bennedsen, Morten, 137
Bertrand, Marianne, 6n, 318, 320n
Besley, Timothy, 239–41
best linear predictor theorem, 37–38
bias: omitted variables, 59–64, 67–68, 116–17, 227; of robust standard error estimates, 294–308; selection, 12–23, 54–55, 64–66, 99–100; of two-stage least squares, 205–17
bias-reduced linearization (BRL), 320
Bindman, Andrew, 12
bivariate probit, 199–205
bivariate regression formulas, 35
Bjorklund, Anders, 148n, 150n
Black, Dan A., 62
Black, Sandra E., 7, 167n26
block bootstrap, 315, 322
Bloom, Howard, 164, 288n
Bloom result, 164–65
Blundell, Richard, 245n
Bond, Stephen, 245n
bootstrap, 300–302; block, 315, 322; wild, 301
Borjas, George, 138n, 194n
Bound, John, 120n, 205n39, 213, 226
BRL. *See* bias-reduced linearization
Bronars, Stephen G., 160
Buchinsky, Moshe, 272, 277
Burgess, Robin, 239–41
Buse, A., 205n39

Cain, Glen G., 53, 59
Cameron, Colin, 321, 322
Campbell, Donald Thomas, 151n
Card, David, 3, 16, 71n, 182, 192, 227, 228–29, 231, 234–36, 241, 243–44, 275n2, 317, 318
Cardell, Nicholas Scott, 189
Carroll, Raymond J., 305n8
causal effect. *See* treatment effects
causality, 3–7; CEF and, 52–53; econometrics and, 113; importance of, 3–4; IV (2SLS) method and, 115–38; and potential outcomes, 13–15, 52–59, 151–58; quantile regression and, 283–86; regression and, 51–68
CEF. *See* conditional expectation function
CEF decomposition property, 32–33
CEF prediction property, 33
censored quantile regression, 275–77
censoring, 102, 275–77
central limit theorem, 43
Chamberlain, Gary, 30n, 41n, 61n16, 68n, 78, 102, 279
check function, 271
Chen, Stacey H., 153n
Chernozhukov, Victor, 213n, 275n3, 277n, 278–79, 283, 284n
Chesher, Andrew, 298, 305n
childbearing and female labor supply, 94, 105–7, 106t, 132–33, 170–72, 172t, 180, 202–5, 203t
chi-square goodness-of-fit statistic, 143
cholera, 227
CIA. *See* conditional independence assumption
classical normal regression model, 47–48

- classical measurement error model, 134n
- class size, and student achievement, 17–23, 19t, 20t, 166n, 263–67, 266t, 315, 316t
- Clements, Nancy, 285n
- clustered standard errors, 312–13, 315, 316t, 320
- clustering, 308–23
- Cochran, William G., 77n
- common support, 76–77, 90–91
- common trends assumption in differences-in-differences models, 230–33, 318
- compliers, 158–161; counting and characterizing, 166–72; in randomized trials, 161–66
- compulsory schooling laws, 117–18, 161, 169t, 170, 182–86, 185f, 193–94
- concentration parameter, 208n
- conditional expectation function (CEF), 29–34; and causality, 52–53; conditional quantile function and, 271; and heteroskedasticity, 46–47, 92–94; regression and, 36–40; in the regression discontinuity (RD) design, 253, 255, 257; and saturated regression models, 49
- conditional independence assumption (CIA), 52–59, 62–63, 69, 176–81
- conditional-on-positive (COP) effects, 98–102
- conditional quantile function (CQF), 270–75, 277, 281
- consistency, 44n8
- constant effects model, 58–64, 115–38, 141–47, 176–78, 253
- continuous mapping theorem, 43–44
- Cook, Thomas D., 151n, 252n, 261n4
- counterfactuals. *See* potential outcomes
- covariates: and causality, 117–21; conditional independence assumption and, 53; in differences-in-differences models, 236–37; exogenous, 120, 126; LATE with, 175–81; and matching, 69–77; and nonlinearity in LDV models, 102–7; and precision, 23–24, 84, 176, 204, 237, 288–89; propensity score and, 80–86; regression and, 23, 35–36; role of, 23; two-stage least squares mistake concerning, 189–90
- CPS. *See* Current Population Survey
- CQF. *See* conditional quantile function
- Crump, Richard K., 90
- Cruz, Luiz M., 215n
- Current Population Survey (CPS), 60, 87–91, 102, 236, 276–77
- Currie, Janet, 148n
- Darwin, Charles, 108
- Davidson, Russell, 209
- DD. *See* differences-in-differences
- Dearden, Lorraine, 63
- Deaton, Angus, 91, 107, 146n
- Dee, Thomas S., 148n, 150n
- defiers, 156–57
- DeGroot, Morris H., 303n
- Dehejia, Rajeev H., 81, 86–87, 243
- delta method, 44
- demand curve, 114, 186–89
- Deming, David, 6
- design effect, 311n

- Devereux, Paul J., 7, 167n26, 210n
deviations from means estimator,
223–24
- Diamond, Alexis, 244n
differences-in-differences (DD),
227–43, 315–19
differencing, 224
discontinuity sample, 256–57,
266–67
- Donald, Stephen G., 314, 321, 322
draft lottery. *See* Vietnam-era draft
lottery
- Duflo, Esther, 318
- Durbin, David L., 230n
- Durbin, James, 114n, 128n
- Dynarski, Susan, 6
- earnings: job training and, 86–91,
88t, 89t, 162–64, 163t, 287–91,
290t; military service and,
63–64, 69–76, 73t, 128–31,
130t, 134–36, 138, 139f. *See*
also returns to schooling; wages
- econometrics, vs. statistics, 113–14
Economic Journal, 109n
economic returns to schooling. *See*
returns to schooling
- educational attainment: compulsory
schooling and, 117–18, 119f,
182–86; family size and, 167
- Education Maintenance Allowance
(Britain), 63
- Education Sciences Reform Act
(2002), 16
- Eicker, Friedhelm,
Eicker-White standard errors. *See*
robust standard errors
- election outcomes, party
incumbency and, 257–59, 258f
- employment: employment
protection, 237–38, 239f; family
size and mothers', 132–33, 133t,
160; minimum wage and,
228–30, 230t, 231f, 233–36,
236t; temporary, 237–38
- endogenous variables, 120, 126–27
- English Poor Laws, 109
- error components, 309
- estimand, 42, 70
- estimator, 42, 117
- Eugenics Society, 109
- Evans, William N., 94, 105, 132,
148n, 150n, 160, 170, 172, 175,
180, 198, 202, 241
- exclusion restriction, 116–17,
153–54
- exogenous covariates, 120, 126
- exogenous variables, 120, 126
- expectation, 30. *See also* conditional
expectation function (CEF)
- experiments: hypothetical, 4; ideal,
4–8; selection bias and, 15. *See*
also random assignment
- external validity, 151, 167
- family size: and educational
attainment, 167; and mothers'
employment, 105–7, 106t,
132–33, 133t, 160–61, 169t,
170, 172, 172t, 180, 202–5,
203t
- Fernandez-Val, Ivan, 275n3,
278–79, 283
- Finn, Jeremy D., 17
- Firpo, Sergio, 283
- first stage, 117–20, 208–9, 212–13
- fixed effects, 23, 221–27, 243–47
- Flores-Lagunes, Alfonso, 209
- forbidden regression, 190–92
42, as ultimate answer, 319
- Freeman, Richard, 224, 225t
- Frisch, Ragnar, 36n
- Frisch-Waugh theorem. *See*
regression anatomy formula

- Frölich, Markus, 283
Frost, Robert, “The Road Not Taken,” 52–53
Fryer, Roland G., 6n
F-statistic, 208, 213, 216, 217–18
- Galton, Francis, 26, 37, 108–9
Gates, Bill, 61n15
Gauss, Carl Friedrich, 109
Gelbach, Jonah, 321, 322
generalized least squares (GLS), 133–37, 197–98, 223n. *See also* weighted least squares
generalized method of moments (GMM), 142
Goldberger, Arthur, 30n, 47, 53, 59
Gosling, Amanda, 283
Graddy, Kathryn, 137, 186
grade repetition, 232–33, 233f
Granger, Clive W. J., 237
Granger causality test, 237–38
Griliches, Zvi, 61n16, 68n, 189, 225n
Grogger, Jeffrey, 160
group-asymptotic approximation, 207n
grouped data, 40, 41f, 92, 133–38, 313–14, 321
Grumbach, Kevin, 12
Guryan, Jonathan, 246
- Haavelmo, Trygve, 7
Hahn, Jinyong, 84–85, 256, 262–63, 267, 277
Hainmueller, Jens, 244n
Hansen, Christian, 213n, 284n, 322, 323
Hansen, Lars Peter, 142, 142n
Hausman, Jerry, 128n, 145, 190, 194n, 210n, 225n
Hausman test, 195n
Head Start preschool program, 11
health insurance. *See* RAND Health Insurance Experiment
Heckman, James J., 76n24, 81, 157, 285n
height, 108
Hernanz, Virginia, 241
heterogeneous effects model, 68–91, 150–88
heteroskedasticity, 45–47, 92–93, 143–44, 274–75, 294–308
heteroskedasticity-consistent standard errors. *See* robust standard errors
Hirano, Keisuke, 8, 83, 85
Hoaglin, David C., 297
Holland, Paul, 14n, 113
Holtz-Eakin, Douglas, 245n10
homoskedasticity, 46, 140, 143, 274, 295–99, 307
Hong, H., 277n
Hopkins, Mark Myron, 189
Horowitz, Joel L., 301
Horvitz, Daniel G., 82
Hotz, V. Joseph, 90
Hoxby, Caroline, 257n
HRT (hormone replacement therapy), 16
Hsia, Judith, 16
Hsu, Li, 307n
- Ichimura, Hidehiko, 76n24, 81
Ichino, Andrea, 83n
identification: in differences-in-differences models, 230; with instrumental variables, 126
identification strategies, 7–8
ILS estimator. *See* indirect least squares (ILS) estimator
Imbens, Guido, 77n, 80n27, 83, 85, 90, 137, 154, 156, 174, 182,

- 186, 210n, 230n, 252, 257n, 284, 286
- incidental parameters problem, 224n4
- incumbency, reelection and, 257–59, 258f
- indirect least squares (ILS) estimator, 121
- individual effects. *See* fixed effects
- Inoue, Atsushi, 149n, 150, 150n
- instrumental variables (IV) method, 113–218; asymptotic inference and, 138–47; and causality, 115–38; constant effects model and, 115–50; defined, 114; fuzzy regression discontinuity and, 259–67; grouped data and, 133–38; heterogeneous potential outcomes and, 150–88; limited dependent variables and, 197–205; origins of, 114; peer effects and, 193–97; quantile treatment effects and, 283–91; randomized trials and, 161–66; two-sample and split-sample IV, 147–50; two-stage least squares, 121–27, 188–92; Wald estimator and, 127–33
- intention-to-treat (ITT) effect, 163
- interaction terms, 37, 49–51
- internal validity, 151
- intraclass correlation coefficients, 310–12
- IV method. *See* instrumental variables (IV) method
- Jackknife IV Estimator (JIVE), 210n
- jackknife variance estimator, 300n
- Jaeger, David, 120n, 205n39, 213
- Jagger, Mick, 61n15
- Jantti, Markus, 148n, 150n
- Jappelli, Tullio, 148n
- Jewitt, Ian, 298
- Jimeno, Juan F., 241
- JIVE. *See* Jackknife IV Estimator
- Job Training Partnership Act (JTPA), 162–64, 163t, 287–91, 290t
- job training programs. *See* Job Training Partnership Act (JTPA); National Supported Work (NSW) program
- Johnson, Norman L., 79
- Johnson, Simon, 4
- JTPA. *See* Job Training Partnership Act
- just-identified, 141, 209, 210, 213
- kappa weighting. *See* Abadie kappa
- Katz, Lawrence F., 34, 274
- Kauermann, Goran, 305n8
- Keane, Dennis, 12
- Kearney, Melissa S., 34, 274
- Kelejian, Harry H., 191n
- Kézdi, Gábor, 318
- Kish, Leslie, 311n
- Kloek, Teun, 309
- Knight, Keith, 43, 44n7
- Koenker, Roger, 272, 275n2, 284
- Kotz, Samuel, 79
- Krueger, Alan B., 7, 17–19, 18n, 20n, 23, 78–79, 115n, 117–20, 123, 126–27, 128, 148–50, 148n, 149n, 153, 153n, 156n, 166n, 170, 183, 207n, 210n, 213–14, 215n, 226–27, 228–29, 231, 309, 317, 318
- Kuersteiner, Guido, 84
- Kugler, Adriana, 241
- labor supply. *See* family size: and mothers' employment
- lagged dependent variables, 243–48

- Lagrange multiplier (LM) test, 145
- Lalonde, Robert J., 16, 87, 243, 252n
- Lang, Kevin, 196, 308, 314, 321, 322
- LATE. *See* local average treatment effect
- latent index model, 97, 157–58, 199–202
- latent variables: probit models and, 97–98; Tobit models and, 100–102
- LATE theorem, 155
- Lavy, Victor, 4, 21, 22n, 23, 160, 167, 256–57, 263–64, 264n, 311, 314n, 321, 322
- law of iterated expectations, 31–32
- law of large numbers, 42, 43
- lead effects, 237–38
- Leamer, Edward E., 78
- least absolute deviations estimator, 271
- Lee, David S., 257–59
- Legendre, Adrien-Marie, 109
- Lemieux, Thomas, 252, 257n, 274, 275n2
- leverage, 297
- Levitt, Steven D., 6n
- Liang, Kung Yee, 312, 319, 323
- limited dependent variables, 94–107, 197–205
- limited information maximum likelihood (LIML) estimator, 209–10, 209–10n, 212, 213
- limiting distributions.
See asymptotic covariance matrix
- LIML estimator. *See* limited information maximum likelihood (LIML) estimator
- linear CEF theorem, 37, 38
- linear location-scale model, 275n2
- linear probability model (LPM), 47, 51, 93, 101–2, 103
- LM test. *See* Lagrange multiplier (LM) test
- local average treatment effect (LATE), 151–58; bivariate probit and, 202; compliers and, 158–61, 166–72; with covariates, 175–81; in fuzzy RD, 262–63; LATE theorem, 155; with multiple instruments, 173–75; and randomized trials, 161–66; with variable treatment intensity, 181–88
- local linear regression, 256
- location shift, 273
- logit, 81, 83, 197, 259n
- Machado, Jose, 282, 283
- Machin, Stephen, 283
- MacKinnon, James G., 209, 298, 299
- Maddala, Gangadharrao, 101n
- Maimonides' rule, 263–64, 265f
- main effects, 37, 48–51
- Mammen, Enno, 301
- Manning, Willard G., 94–95
- Manski, Charles F., 30n
- marginal effects, 80, 103–7, 197, 204
- marginal quantiles, 281–83
- Mariano, Roberto S., 209
- Mark, Steven D., 82
- Mason, William M., 61n16, 68n, 189
- Mata, Jose, 282, 283
- matching: benefits of, 69; common support and, 76–77, 90–91; drawbacks of, 57; and the propensity score, 80–91; regression vs., 69–80

- McCaffrey, Donald F., 320–21, 320n
- McClellan, Mark B., 170
- McCrary, Justin, 259
- McDonald, John F., 101
- McNeil, Barbara J., 170
- MDVE. *See* Minneapolis Domestic Violence Experiment
- Medicaid, 242–43
- Meghir, Costas, 283
- Melly, Blaise, 283
- Meltzer, Allan, 27–28
- Messer, Karen, 300n
- Meyer, Bruce D., 230n, 241
- Michaels, Guy, 137
- Milgram, Stanley, 5
- military service and earnings, 63–64, 69–76, 73t, 128–30, 130t, 134–36, 139f, 153, 168, 169t
- Miller, Douglas L., 321, 322
- minimum mean squared error (MMSE), 33
- minimum wage, employment effects of, 228–31, 230t, 232f, 233–37, 236t
- Minneapolis Domestic Violence Experiment (MDVE), 165–66
- Mitnik, Oscar A., 90
- MMSE. *See* minimum mean squared error
- Moffitt, Robert A., 101, 242
- monotonicity, 154–58
- Monte Carlo results: for instrumental variables (IV) estimators, 210–12, 211f, 212f; for robust standard errors, 304–8, 306t
- Moreira, Marcelo J., 215n
- Morgan, Mary S., 115n
- Moulton, Brent, 294, 310
- Moulton factor, 308–15, 322, 323–25
- Moulton problem, 294, 312–15, 319–23
- Mullainathan, Sendhil, 6n, 318
- multiple instruments, LATE with, 173–75. *See also* overidentification
- National Health Interview Survey (NHIS), 12–13
- National Longitudinal Survey of Youth (NLSY), 61, 62t, 85
- National Merit Scholarships, 252
- National Supported Work (NSW) program, 87–91, 88t, 89t
- natural experiments, 7–8, 21
- Nelson, Charles R., 205n39
- Neumark, David, 235n
- never-takers, 158–61
- Newey, Whitney K., 82, 138n, 142n, 146n, 191, 245n
- Newhouse, Joseph P., 170
- NHIS. *See* National Health Interview Survey
- Nickell, Stephen, 245
- NLSY. *See* National Longitudinal Survey of Youth
- nonlinearity, 77–80, 102–7, 181–88, 190–92, 253–57
- NSW. *See* National Supported Work (NSW) program
- Nurses Health Study, 16
- Obenauer, Marie, 228n
- observational data, regression and, 29
- Occam's razor, 107
- OLS. *See* ordinary least squares
- omitted variables bias (OVB), 59–64, 67–68, 116–17, 227; formula, 60

- ordinary least squares (OLS), 34–51;
asymptotic inference, 40–48; vs.
limited dependent variables
models, 94–107, 197–98; and
marginal effects, 105, 107, 197;
vs. matching, 69–80; vs.
propensity score methods,
80–91; vs. random effects,
223n; vs. weighted least squares,
91–94
- Oreopoulos, Philip, 160
- Orr, Larry L., 16
- OVB. *See* omitted variables bias
- overidentification, 141–47, 166,
209, 213
- overidentification test, 143–47, 166
- panel data, 222–27, 243–46
- Panel Study of Income Dynamics
(PSID), 223
- participation effects, 98
- Pearson, Karl, 109
- peer effects, 193–97
- Perry preschool project, 11, 12n
- Pfefferman, Daniel, 91
- Pischke, Jörn-Steffen, 109n, 148n,
196, 232
- pivotal statistic, 302
- Porter, Jack, 256
- Portnoy, Stephen, 275n2
- potential outcomes, 13–15; and
differences-in-differences,
228–29; and fixed effects,
222–23; and instrumental
variables, 115–17, 150–58; and
matching or regression, 52–59;
and quantile treatment effects,
284–85; and regression
discontinuity designs, 253–55
- Poterba, James, 230n
- poverty rates, 109
- Powell, James L., 107, 276
- prescreening, 90–91
- probit model, 97–98, 104–7;
bivariate 199–205
- propensity score methods: common
support and, 90–91; control for
covariates using, 80–86; full
covariate matching vs., 83–86;
and multivalued treatments,
80n27; regression vs., 82–83,
86–91; and weighting, 82–83
- propensity score theorem, 80–81
- proxy control, 66–68
- PSID. *See* Panel Study of Income
Dynamics
- quantile regression, 269–83;
approximation property of,
277–81; and causality, 283–84;
censored, 275–77; marginal
quantiles, 281–83; standard
errors for, 275n3; and treatment
effects, 283–91
- quantile regression approximation
theorem, 278
- quarter-of-birth instruments,
117–20, 119f, 123–26, 124t,
128, 129t, 166, 169t, 170, 183,
213–15, 214t
- quasi-experiments, randomized
trials vs., 21
- RAND Health Insurance
Experiment (HIE), 94–101, 96t
- random assignment: as benchmark
for causal inference, 4–7, 11–12,
21; importance of, 11–12;
instrumental variables in,
158–59, 161–66; practicality
of, 19–21; and selection
bias, 15; in social science
experiments, 16; success of, 18;
value of, 15–22

- random effects, 223n, 309
- randomized trials. *See* experiments; random assignment
- RD. *See* regression discontinuity
- reduced form, 117–20, 213
- regression, 22–24, 27–110; and bad controls, 64–68; and causality, 51–68; and the conditional expectation function, 34–40; conditional independence assumption and, 52–59; differences-in-differences and, 233–41; with fixed effects, 23, 221–27; fundamentals of, 28–51; with limited dependent variables, 94–107; matching and, 69–91; and observational data, 29; omitted variables bias formula and, 59–64; propensity score methods vs., 86–91; weighted vs. unweighted, 91–94. *See also* quantile regression; regression discontinuity
- regression anatomy formula, 35–36
- regression CEF theorem, 38, 49
- regression discontinuity (RD), 251–67; fuzzy, 259–67; sharp, 251–59
- regression to the mean, 108–9
- research agenda: causal relationship of interest in, 3–4; ideal experiments and, 4–7; and identification strategy, 7–8; and statistical inference, 8
- returns to schooling, 3–4, 52–62; ability bias and, 60–62, 66–68; bad control and, 64–68; censoring and, 102; conditional expectation function and, 29–31, 31f, 39f; conditional-on-positive effects and, 102n; IV (2SLS, Wald) estimates of, 115–21, 119f, 123–26, 124t, 128, 129t, 161, 169t, 182–86, 185f, 192–95, 213–15, 214t; omitted variables bias formula and, 59–62, 62t; quantile regression estimates of, 272–75, 273t, 280f; twins estimates of, 226–27
- Richard, Scott, 27–28
- Ridder, Geert, 83, 85
- Robins, James M., 82
- Robinson, James A., 4
- robust standard errors, 45–48, 293–308
- Rosen, Harvey S., 245n10
- Rosenbaum, Dan T., 241
- Rosenbaum, Paul R., 77n, 80, 84
- Rosenzweig, Mark R., 132
- Roth, Philip, *The Human Stain*, 6
- Rouse, Cecilia, 226
- Rubin, Donald B., 8, 14n, 71n, 77n, 80, 84, 156
- Rubin causal model, 14n. *See also* potential outcomes
- Ruud, Paul A., 80n26
- Salvanes, Kjell G., 7, 167n26
- sample selection model, 101n
- SAS, 46, 92, 122, 188, 209n
- saturate and weight, 177–78
- saturated regression models, 37, 48–51, 234
- Schervish, Mark J., 303n
- Schlosser, Analia, 160, 167
- schooling. *See* family size; peer effects; returns to schooling
- school start age: and educational attainment, 117–20, 119f; identifying effects of, 6–7. *See also* quarter-of-birth instruments
- school term length, student achievement and, 232–33, 233f

- Segal, Lewis M., 93n
selection bias, 12–24; bad control and, 64–66; conditional independence assumption and, 54–56; conditional-on-positive effects and, 99–100
selection on observables. *See* conditional independence assumption; regression discontinuity (RD): sharp
SEM. *See* simultaneous equations model
serial correlation, 294, 315–19
Shadish, William R., 151n
Shakespeare, William,
As You Like It, 6
Shatner, William, 5n
Shore-Sheppard, Lara, 312
sibling sex composition, 132–33, 133t, 169t, 170, 171–72, 172t, 202–5, 203t
simultaneous equations model (SEM), 114–15, 120–21, 126–27
skewed distribution of outcomes, 101
Slutsky’s theorem, 43–45
Smith, Jeffrey, 76n24, 87n, 285n
Snow, John, 227
social return to schooling, 193–95
Solon, Gary, 149n, 150, 150n, 226
Somebody Else’s Problem, 315–16, 316n
Souleles, Nicholas S., 148n
split-sample instrument variables, 147–50
standard errors: asymptotic OLS, 40–48; asymptotic 2SLS, 138–40; bias of robust, 294–308; bootstrap, 300–302, 315, 322; for class size effects, 316t; clustered, 312–13, 320; conventional vs. robust, 307; for data with a group structure, 308–15, 319–23; for LIML, 209n; Monte Carlo results for, 306t; Moulton, 312, 322; for panels with serial correlation, 315–19; for quantile regression, 275n3; robust, 45–48, 293–308; sampling variability of, 305–8
Stanley, Julian C., 151n
STAR experiment. *See* Tennessee STAR experiment
Startz, Richard, 205n39
Stata, 40, 45, 46, 91, 92, 104, 107, 122, 188, 209n, 287, 293, 294, 312n13, 315, 319, 323
statistical software packages. *See* SAS; Stata
statistics, vs. econometrics, 113–14
Stigler, George, 110
Stigler, Steven, 109
Stock, James H., 115n, 213
structural equation, 121
student achievement: class size and, 17–24, 19t, 20t, 166n, 263–67, 266t; National Merit Scholarships, 252; school start age and, 6–7; school term length and, 232–33, 233f
Sullivan, Daniel, 71n
Taubman, Paul, 61n16
temporary workers, 237–38, 239f
Tennessee STAR experiment, 17–24, 166n, 266, 267, 309, 312, 315, 316t
Thompson, Donovan J., 82
timing, causal relationships and, 68, 237–38
Tobit model, 100–7, 106t
Todd, Petra E., 76n24, 81, 87n, 256, 262–63, 267

- TOT. *See* treatment effects: average effect of treatment on the treated
- treatment effects, 3–4, 13–15, 22, 54–57; average effect of treatment at a covariate value, 54–57; average effect of treatment on the treated (TOT) 14, 54, 70–71, 76, 81–83; and limited dependent variables, 94–107, 197–205; quantile, 283–91; regression discontinuity and, 252–59; selection bias and, 13–15; unconditional average (ATE), 15, 72, 82–83. *See also* constant effects model; heterogeneous effects model; local average treatment effect (LATE)
- Trebbi, Francesco, 115n
- Trochim, William, 260n
- twins: estimates of returns to schooling using, 226–27; as instruments for family size, 132–33, 133t, 160–61, 169t, 170, 172, 172t
- two-sample instrumental variables, 147–50
- two-stage least squares minimand, 141–47
- two-stage least squares (2SLS), 121–27; asymptotic inference, 138–47; and covariate ambivalence, 189–90; forbidden regressions, 190–92; manual procedure for, 188–89; minimand 141–47; mistakes, 188–92; two-sample, 147–50. *See also* instrumental variables (IV) method
- union membership, wages and, 221–25, 225t
- unit causal response, 183
- validity: external, 151, 167; internal, 151
- Van der Klaauw, Wilbur, 256, 261–63, 261n5, 267
- variable treatment intensity, 77–80, 181–88. *See also* nonlinearity
- Venti, Steven, 230n
- Vietnam-era draft lottery, 128–31, 130t, 134–38, 168, 169t
- Viscusi, W. Kip, 230n
- visual IV (VIV), 138, 139f, 141, 144
- von der Nienburg, Bertha, 228n
- wages: distribution of, 269–70, 272–75, 273t; union membership and, 221–25, 225t. *See also* earnings; minimum wage; returns to schooling
- Wahba, Sadek, 81, 86–87, 243
- Wald, Abraham, 114n, 128n, 145n
- Wald estimator, 127–38, 129t, 130t, 133t, 145n, 146, 183
- Wald test, 145, 145n
- Wascher, William, 235n
- Waugh, Frederick V., 36n
- weak instruments, 205–16
- weighted average of unit causal effects, 181
- weighted least squares (WLS), 39–40, 91–94, 136–37, 149, 180, 256, 313–14, 321
- weighting: grouped data and, 39–40, 41f, 92, 136, 144, 149, 313–14, 321; heteroskedasticity and, 92–94; matching and, 71–77; propensity score, 82–83; regression, 91–94; saturate and weight, 177–78; with Abadie kappa, 179–81, 203–5, 286–87

- welfare benefits and migration, 241–42
- Welsch, Roy E., 297
- West, Kenneth D., 142n, 146n
- White, Halbert, 45, 46, 140, 142, 143, 298, 299, 300n, 304, 305, 312
- White standard errors. *See* robust standard errors
- wild bootstrap, 301
- Wise, David, 230n
- within estimator, 224
- WLS. *See* weighted least squares
- Wolpin, Kenneth I., 132, 160
- Wong, Vivian C., 252n
- Wooldridge, Jeffrey, 104, 222n, 223n, 309n
- Wright, Jonathan H., 213
- Wright, Phillip, 114
- Wright, Sewall, 114
- Yang, Song, 307n
- Yelowitz, Aaron, 148n, 242
- Yitzhaki, Sholomo, 78
- Yogo, Motohiro, 213
- Yule, George Udny, 109
- Zeger, Scott L., 312, 319, 323
- Zhao, Lueping, 307n