Conclusion

So ends our two books.

You have probably noticed that my theses—that economics' central problem is credibility, and that this devolves from inattention to craftsmanship—are, in the terms of that old Scottish verdict, not proven. You've got me there. My examples are varied, in almost every way, but you can't prove something on examples alone, and examples is all I have.

But these theses aren't merely retrospective, and need not be tested only by me. Test them yourself, prospectively, as you attend conferences, watch seminars, read papers, and write your own. If they hold true, you will confront the issues raised in this book at every turn, and will see how the principles I've laid out can ameliorate them.

If you begin by applying these principles back to this book itself, you will realize that I have one imperative remaining: to close the loop, by articulating this book's methodological organizing principle and comparing it with the alternatives.

No depiction of these alternatives can be fully just. None of them have been formally stated or justified, except perhaps for specific research questions; none has an official spokesperson or body of arbiters. Still, who are we kidding? We all can recognize three distinct strains of thought, three visions of the research process so well known that they mold the profession's self-image: what it is that economists are supposed to do when they do economics. These visions (and their primary exponents) have been represented in the preceding chapters. Let's take them out and hold them up to the light.

The first vision emphasizes one pillar of the profession: theory. Any model not built upon economic bedrock is problematic for policy. Only structural models founded on behavioral and technological primitives can provide the true economic understanding that is needed to fix economic problems. While this organizing prin-

D. Grant, *Methods of Economic Research*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-01734-7

ciple has been most forcefully expressed in macroeconomics, its spirit survives in micro as well, in some of the more elaborate structural models that arise in a variety of fields.

The danger in this approach is that it downplays the tradeoffs that are intrinsic to modeling. The fact that causality is a desirable quality of models doesn't make the reach/grasp problem disappear, nor any infirmities of your data, etc. If we bury the process at work under an avalanche of formalism, we risk developing models that are impractical to estimate and lacking in versatility.

When this happens, the easiest way to cover your tracks is by sandbagging testing. What null should the model tested against? How should it be compared to competing models? Which elements of the system should be included in these tests? How "good" must the evidence be to pass the threshold of policy usefulness? Simply testing your main parameter estimates against the null hypothesis of zero, or variations on that theme, ignores all of this. Then the conclusions we reach may be strong, but not credible in the end.

By emphasizing one input of the research process to the detriment of others, this vision does not sufficiently distinguish a theory from a model. In doing so, it diverts us from the ultimate obligation of modeling: to ensure, with everything we have—description, vernacular knowledge, hypothesis testing, closing the loop, etc.—that we provide a description of process and outcome that can be taken seriously.

A second vision emphasizes the other pillar: estimation. Garden-variety econometric estimates are suspect, because there are so many unanticipated or unfixable sources of bias. Only estimates based on genuine natural experiments, involving randomization, discontinuities, matching, etc., give us the "clean identification" we need to ensure our estimates are sound.

This approach isn't foolproof, as we saw in Chap. 8's discussion of Goldin and Rouse (2000) and Angrist and Krueger (1991). To follow the whole tick-tock on the latter paper—the original, "small-sample" criticism by Bound et al., Angrist's response in *Mostly Harmless*, the counterevidence in the working paper version of Buckles and Hungerman (2013), Angrist's response in *Mastering Metrics*, the more definitive counterevidence in Buckles and Hungerman's final paper—is to despair whether economists will ever agree on anything. Finally I dug up the two "pre-existing" papers, downplayed by the original and by Buckles and Hungerman, which addressed the central issue of the independence of birth quarters (Kestenbaum 1987; Lam and Miron 1991, available earlier as a working paper). Mmmmm. Angrist and Krueger never should have started down this road to begin with.

This vision, like the first, shunts some inputs into the research process off to the side. In doing so, it gives us a clean causal estimate, under suitable circumstances, but often obscures our broader understanding what is really going on.[1] Even then, craftsmanship is not rendered irrelevant. It is needed to determine whether the circumstances are suitable to begin with, as we just saw. It is needed even more when

[1] Commentary by Leamer, Keane, Sims, and Nevo and Whinston in the Spring 2010 *Journal of Economic Perspectives* nicely fleshes out this point and its implications.

no clean natural experiment is available, a situation for which this vision offers much less guidance. Most important economic questions fall into this category.

The third vision, put forward by Milton Friedman (1953), was spawned by consternation over the fact that managers rarely sounded like they followed MR = MC when asked (Hall and Hitch 1939). It emphasizes neither pillar per se, nor any other input into the research process. A good model is one that consistently makes good out-of-sample predictions, period.

The vision would seem to accord with Chap. 10's emphasis on ends, not means. After all, if a model has fantastic out-of-sample predictive power, who really cares how it was constructed? The problem with this sentiment is that it is a theoretical proposition rather than an empirical one. Had Friedman followed his essay with a companion piece titled "I Told You So," filled with literature-based examples of nonsensical, uncontrived, highly-predictive economic models, that would be one thing. There is no such paper, and there never will be.

Did Friedman foresee that, by limiting the evaluation criterion to one largely unreachable, untestable goal, his essay would become the intellectual foundation for every half-baked analysis that failed to seriously consider the soundness of its model, data, and results? Did he anticipate that, for many policy decisions, $Y = ßX + ε$ would be a major improvement over the alternative, despite its faults? Or were these out-of-sample predictions he got wrong?

Shoot. If the deal is that people can't always articulate why they do things, you needn't base your counterargument on the philosophy of science. All you need is this (Scott 1998, p. 329):

> Some knowledge is so implicit and automatic that its bearer is at a loss to explain it. A staple of early medical training, I have been told, is the story of a turn-of-the-century physician who was fantastic at diagnosing syphilis in its early stages. Lab tests confirmed his diagnoses, but he himself did not know just what it was that he detected in his examinations that led him to his conclusions. Intrigued by his success, hospital administrators had two other doctors closely observe his examinations over several weeks. At long last, they eventually realized that he was unconsciously registering a slight eye tremor in these patients. These tremors then became a universally recognized symptom of syphilis.

As an aspiration, each of these visions is admirable. We all want fully causal, impeccably-identified economic models with outstanding predictive power. In practice, however, each vision emphasizes one aspect of the research process to the diminution of others, thereby engaging in a touch of what Thomas Mayer (1993, p. 57) derisively calls "The Principle of the Strongest Link." The stress placed on causality, identification, or model versatility distorts the tradeoffs that are involved in modeling, diverts our attention away from other elements of the research process, which is generally quite multifaceted, and offers limited guidance in the many situations in which these visions are untenable.

This would pose little problem if research elements effectively substituted for each other, so that excellence in one compensated for shortfalls elsewhere. This is not the testimony of experience. Good theory will not compensate for lousy data, weak testing, or a poor understanding of the situation being modeled, and neither will anything else. In general, the opposite is true: these research elements are

complementary, as I have claimed repeatedly throughout this book. They form a system.

This fact implies a different vision of the research process: as a confederation of fallible elements, an assembly of many potentially faulty components, each of which depends on the others to work. The study of such assemblies falls under what systems engineers would call "reliability theory." This theory undergirds Gavrilov and Gavrilova's mortality model in Chap. 3, which treats the human body as a corporation of components, each of which is made from imperfect parts, and each of which must function for the organism to succeed.

Gavrilov and Gavrilova point out that system reliability can be achieved in two ways. One way is to construct each component from quality parts. This is common in the mechanical realm, say in producing stereo systems, because each part can be sturdily made before its assembly into the component, and each component carefully tested before its assembly into the final product. The other way is to have redundancy in the parts that each component is composed of. Then the failure of one part is inconsequential so long as the others continue to function. This is common in the biological realm, where sturdiness cannot be assured and testing is impossible.

The first way of achieving reliability is akin to the pure application of the scientific method. Humans have the power to control each element of this research process: theory, experimental design and execution, data collection, and hypothesis testing. Accordingly, it is reasonable (though somewhat Pollyannish) to visualize "pure science" as a clean, clear, trustworthy sequence of procedures.

Social scientists rarely have this power. We cannot place too much faith in any one element of our research. Most of the time, we can't control the variables that we measure and how well we measure them, can't design our own experiments and eliminate confounders or alternative explanations from consideration, can't fully understand how our findings might depend on context or institutional detail. Then to think of the research process in the first way is to labor under a mistake. Our approach to achieving reliab—I mean, credibility—must be more like the second way, which is suited to a world of weakness, not strength.

In a system of complementary elements, we must attend more to those that are weaker, not stronger—and with limited control over the quality of each element, everything is suspect. The implications of this research vision have resounded throughout this book. Respect this weakness by employing every research element at your disposal, not just the two pillars of the profession, and by honoring the principles of self-determination and seeing the problem on its terms. Compensate for possible flaws in any one facet of your research by embracing the principle of redundancy and its cousin, transparency. Acknowledge the complementarities between research elements via the principle of continuity, which de-compartmentalizes the research process and strengthens the connections between these elements.

This weakness extends to the model's evaluation and application, which occur in a world with many possible influences on behavior, both theoretical and contextual. This fact makes model versatility somewhat elusive and traditional hypothesis

testing somewhat effete. Accordingly, embrace the reach/grasp problem and the tradeoffs therein, and base your model on a thorough, well-rounded understanding of the phenomenon of interest, achieved using vernacular knowledge and description. Broaden hypothesis testing in the ways we have discussed, and examine your model's collateral implications by closing the loop. Adhere to the principle of harmony, since amassed evidence that a model captures the essence of the process at work is an important sign of its fidelity and versatility. All of this is the opposite of genius—and the essence of craftsmanship.

I get it. For all of this you are guaranteed nothing in the end. Things don't have to work out. Throw everything you have at Nigeria's Child Rights Act, and you will still come away knowing little about its effect on child marriage. The system is too large and sluggish, the experimental content too limited for econometrics of any sort to be decisive. But there are plenty of cases in which craftsmanship carries you a long way. You may even achieve that ideal of coherence, in which everything falls into place, like the carnival problem in the previous chapter.

To anyone who believes otherwise, who argues that these weaknesses are so daunting that the whole exercise can't really be taken seriously, I say: poppycock. You wouldn't be saying this if you've ever achieved anything like coherence, and if you've never done that, then how can you know? Your lack of experience needn't extend to everyone. This view of social science is more social than science.

Timidity like this, an atavistic fear of being so fully exposed, is what makes this vision of applied microeconomic research seem a little radical. It is tempting to remain where we are, sheltered by excessive formalism, uncertain experimental content, and ambiguous language. For this reason, implementing some of the principles articulated in this book requires a certain independence, a willingness to stand aside from the rushing current.

This independence is, I recognize, a genuine barrier to these ideas' adoption, particularly for readers only newly acquainted with the profession. The practical benefits of lying securely within the mainstream are many: it flatters the work of potential referees, reduces or eliminates methodological confusion, and avoids raising the hackles of tepid reviewers who, unsure of their ability to assess things independently, view anything non-routine, however sensible, as suspect. These concerns are not irrational. So, the thinking goes, go along to get along. You're just doing your job—you and everyone else.

In response, all I can tell you is this: be brave. Luminous beings are we! The conventions of the world are not as mighty as we think, and the courage required to confront them not theatrical, like Mel Gibson in *Braveheart*, but something quieter and more true. So let out your Tookish side. I think you will gain something in the end.

I wrote this book for a reason we all share—a latent discontent that develops early in our training, a queasy feeling that something is just not right. I know that feeling. It gnawed at me for 25 years. I feel better now.

References

Angrist J, Krueger A (1991) Does compulsory school attendance affect schooling and earnings? Q J Econ 106(4):979–1014

Buckles KS, Hungerman DM (2013) Season of birth and later outcomes: old questions, new answers. Rev Econ Stat 95(3):711–724

Friedman M (1953) The methodology of positive economics. In: Friedman M (ed) Essays in positive economics. University of Chicago Press, Chicago

Goldin C, Rouse C (2000) Orchestrating impartiality: the impact of blind auditions on female musicians. Am Econ Rev 90(4):715–741

Hall R, Hitch C (1939) Price theory and business behavior. Oxf Econ Pap 2:12–45

Kestenbaum B (1987) Seasonality of birth: two findings from the decennial census. Soc Biol 34(3–4):244–248

Lam D, Miron J (1991) Seasonality of births in human populations. Soc Biol 38(1–2):51–78

Mayer T (1993) Truth vs. precision in economics. Edward Elgar, Brookfield, VT

Scott JC (1998) Seeing like a state: how certain schemes to improve the human condition have failed. Yale University Press, New Haven

Glossary

Abjure Renounce; give up entirely.
Accoutrements Embellishments that serve a purpose, rather than being merely frivolous.
Accretion A slow accumulation, built up layer upon layer like a pearl.
Ad hoc Created for an immediate purpose only, with no larger goal in mind.
Aficionado Someone who is passionate and knowledgeable about some subject or activity.
Anoxic Without oxygen.
Apotheosis The ideal achievement or level of achievement.
Armamentarium All the stuff that is available for you to use in performing a duty or accomplishing a task.
Atavistic Raw and primitive.
Autoimmune disease A disease in which someone's immune system turns on that person and makes them sick.
Ballpark As a noun, it means "to be in the broad neighborhood of"; as a verb, it means to make an approximation that is intended to be in that neighborhood.
Beef Problem or objection.
"Between a rock and a hard place" A phrase implying a situation in which there are no good options, only various unsatisfactory options to choose from.
Body shop A place that fixes non-mechanical parts of a car, such as the frame or the exterior.
Borax A mineral containing boron that is better used in laundry detergent than in food.
Bucolic Pastoral; very rural.
Cadastral Having to do with land surveys that establish property boundaries.
Catholic Universal or wide-ranging.
CFA, or CFA franc Roughly speaking, a currency used by several West African countries.
Chestnut A real classic from the past.
"Come up sevens" A rare event in slots (gambling) in which you win big.

D. Grant, *Methods of Economic Research*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-01734-7

Communitarian A social structure that revolves around each individual contributing to the community.
Conflux Coming together.
Consonance Coming together in concurrence or harmony.
Consumption germs Tuberculosis (which, ironically, also follows a random walk).
Conundrum A mystery wrapped in an enigma stuffed inside a puzzle; the turducken of problematic situations.
Coriolis Effect When something is moving within a rotating medium, such as the Earth, this force pushes the object perpendicular to its direction of motion.
Countermand To dictate a reversal of a previous order or act.
Cutting room floor Refers to the editing of movies that are shot on film; material that was excluded from the final version of the movie ended up on the cutting room floor.
De rigueur Something the elite, especially, are socially expected to do.
Deep into the weeds Really getting into details.
Dicta Plural of dictum, which is like a very strong dictate.
Dot-matrix An old type of printer, common in the 1970s and 1980s, that printed everything using little dots that were aligned on a grid.
Dyad A connection between two points on a graph; more generally any pair of like things.
Effectuates Makes happen.
Effete Worn out; exhausted in substance or content.
Electorate Potential (but not necessarily actual) voters.
Escarpment Like a long cliff.
Exigencies Things that must be urgently done.
Exoskeleton A skeleton that is on the outside, like lobsters or scorpions have.
Exponents People who advocate for an idea.
"A feature, not a bug" What we want to happen, not what we don't.
Fixtures A term used outside the U.S. for the schedule of games for a sports team or league.
Flyover country A pretentious phrase used by some coastal Americans to describe people living in the interior of the country.
"For crying out loud" An expression of great frustration.
Fortran An early computer programming language that was used especially for mathematical tasks.
Fractal An often-complicated pattern that repeats itself on increasing spatial scales.
Functional fixedness Getting so absorbed in the details that you lose the big picture.
G-7 countries A group of seven major industrialized countries, including the U.S., U.K., France, Germany, and Japan.
Garden-variety Ordinary; typical; unexceptional.
GED A substitute for a regular high school diploma that you get by passing a series of tests on English, math, etc. While I was in college, both of my parents studied for and obtained their GED's.

Glycerin A compound that is or has been used in antifreeze, electronic cigarettes, and food.

"Grist for the mill" Something substantive to work on.

Guanxi Who you know, or a term for the social capital possessed by knowing someone of value to you, such as a doctor.

Gumbo A soup, traditional in South Louisiana, that included whatever meat and seafood the cook had left over, including rabbit, crawfish, shrimp, sausage, and chicken.

Haka A cool, ritual chant and dance that originated among the Maori in New Zealand and is becoming more popular around the globe.

HAL The main supercomputer in the movie *2001: A Space Odyssey*, by Quentin Tarantino.

"Hang your hat" Place your confidence in; rest your case on.

"Hard-up" Poor, in poor shape, or both.

Hoary Really, really old, almost too old.

Homophily Like likes like.

Huckster An aggressive pitchman for a product or service.

Hydra A multi-headed monster from Greek mythology, which grows two new heads every time one is lopped off.

"In spades" In large, generous quantities or amounts.

Inured To get so used to something that you hardly even notice it anymore.

Ironclad Absolutely certain.

Keiretsu A network of businesses that are linked together, financially and otherwise.

Kinesthetic Active; dynamic; full of movement.

Laterite A mineral-laden, often-clayish soil that is hard to grow things in.

"Levers to press" In the archetypal (yet cartoonish) Taylorist factory, you have a set of levers and pressing the right one gives you the result you want.

"Low-hanging fruit" The fruit that is easiest to pick; opportunities that are easy to seize.

Likert Scale A multi-option, graduated survey response scale, e.g., "strongly agree," "somewhat agree," … "strongly disagree."

Manifest variable In psychology, a variable that is actually measured (as opposed to a latent variable, which isn't).

Marlin-spike A tool that sort of looks like a big needle, which is used to work with rope on ships even today.

Missive A letter, often long, and often official.

Nantucket A long, thin island off the Massachusetts coast that used to be a haven for whaling vessels, and now has vacation homes for rich people.

Neuroticism General nervousness and anxiety.

Non-sequitur Two statements, the second of which does not follow from the first.

"Not even wrong" Refers to Wolfgang Pauli's derogation of theories that make no testable predictions, and thus which cannot be proven wrong.

Obfuscating Covering things up or deliberately confusing things.

"Oh snap" A somewhat outdated phrase that indicates that one person successfully made fun of another.

"Old-school" How it used to be done in the somewhat-distant past.
"On point" Right on the mark.
"One-off" An isolated case or unique event.
Orogeny Mountain-making.
Pandora's Box In Greek mythology, a box that, when opened, unleashed all kinds of mayhem.
Pantheon A literal or figurative collection of highly honored people (or their remains).
Pejorative Derogative; putting people down.
Pernicious Particularly bad or evil.
Piggy banks Hollow ceramic animals with slits in the top, in which children put their allowance or spare change in order to save up for college, like I did.
Pollyannish Overly and unrealistically optimistic.
Pong The first video game, basically table tennis on a computer.
"Poppycock" I don't think so.
Potpourri A fragrant, colorful mixture composed of pretty leaves, flower petals, etc.
Preponderance of the evidence A legal standard, in contrast to "beyond a reasonable doubt," that gives victory to whichever side the evidence most favors.
Prescient Accurately foresaw the future.
"Primordial ooze" Refers to the "soup" that life is said to have first originated in.
Quiver The thing that archers carry on their back that holds arrows.
Quotidian Boring, everyday duties that somebody has to do.
"A reasonable way to identify pornography" Refers to Justice Stewart's extremely informal test for obscenity in a famed Supreme Court case.
Re-christening This is what happens when a ship is given a new name.
Rube Goldberg machine A complicated contraption that accomplishes a simple task, named after the cartoonist who invented them.
Rule of reason A legal standard in antitrust law that, in contrast to a per se standard, limits intervention only to those actions that are, on balance, sufficiently anticompetitive.
Sallies Gentle thrusts or outward movements.
Sandbagging Deliberately underperforming, due to lack of effort or diligence.
School accountability movement A political movement, spearheaded in the American South, to assess how well schools are performing through their students' scores on standardized tests.
Secondary producer A producer of recycled metal.
Senescence Old as f—.
Sextant This odd-looking tool seafarers use in navigation.
"Shootin' distance" Pretty close but not too close.
Standardized tests Tests created by big companies used not for class grades but to determine how much a student knows in some more general sense.
"Swing through the ball" In sports, the act of completing the swing of a bat, golf club, etc., through and beyond its contact with the ball.
Synoptic Big-picture; allowing a general overview of the subject.

Systems engineers Just what it sounds like; what operations management would be if it focused on planning and design instead.

Tapestry A beautiful, elegant cloth, woven from threads of many colors, that royalty of old might have hung on the wall.

Tell A behavior or body movement that signifies something about the cards that a gambler is holding.

Terrigenous Just like it says in the text: sediments from the land going into the sea.

Thereunto appertaining A pompous phrase used in college graduations. I have no idea what it means.

"Tick-tock" Sequence of events, told as if in real time.

"To boot" In addition.

Tookish Adventurous beyond people would expect of you.

Transom A (typically) flat surface that forms the very back of a ship.

Trifecta A bet involving three predictions, which all must be correct to win.

Typification Letting the type represent the whole, as if Pepe the Frog and Kermit the Frog were equivalent because they are both frogs.

Unostentatiously Without drawing attention to itself (which, ironically, this word fails to do).

Variegated Something that is highly mottled in color or, by analogy, that contains a mixed variety of ingredients.

Vignette A little story, more brief than an anecdote.

Welter A disorderly mix.

"Window dressing" Something that looks nice but is not functional; something merely for appearance.

Index

D. Grant, *Methods of Economic Research*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-01734-7