

Contents

Preface | xix

PART I. EXPERIMENTAL METHODS IN ECONOMICS

1. Introduction | 3

Key Ideas | 3

1.1 Causal Inference | 4

Experimental Problem 1: Quantifying Economic Fundamentals,
Measuring Treatment Effects, and Identifying Key Mediators and
Moderators in an Ethically Responsible Manner | 5

Experimental Problem 2: Predicting Whether the Causal
Impacts of Treatments from One Environment Transfer to
Spatially, Temporally, Scale-, or Population-Differentiated
Environments | 6

1.2 The Book's Game Plan | 7

Notes | 10

References | 10

2. A Primer on Economic Experiments | 13

Key Ideas | 13

2.1 Four Running Examples | 15

2.2 The Empirical Approach in Economics | 16

2.3 Experiments in Economics | 21

2.3.1 Laboratory Experiments | 21

2.3.2 Field Experiments | 25

2.3.2.1 Seven Criteria That Define Field Experiments | 26

Experimental Subjects: Population and Selection | 27

Experimental Environment | 28

2.3.3 What Parameters Do the Various Experimental Types Recover? | 31

- 2.4 What Experimental Type to Choose | 36
 - 2.4.1 Control across the Experimental Spectrum for Identification Purposes | 37
 - 2.4.2 Control across the Experimental Spectrum for Measurement Purposes | 37
 - 2.4.3 The Ability to Replicate across the Experimental Spectrum | 38
 - 2.4.4 Control across the Experimental Spectrum for Inferential Purposes | 39
 - 2.4.5 Control across the Experimental Spectrum to Ensure External Validity | 42
- 2.5 Conclusions: Key Complementarities Exist across the Lab and Field | 44

Appendix 2.1: Introducing General Potential Outcomes Notation | 46

Notes | 48

References | 50

3. Internal Validity: Identification in Economic Experiments | 55

Key Ideas | 55

- 3.1 Four Running Examples | 56
- 3.2 The Assignment Mechanism | 58
- 3.3 Potential Outcomes Framework | 60
- 3.4 From Individual Treatment Effects to Average Treatment Effects | 63
- 3.5 How Selection Leads to Bias | 67
 - 3.5.1 Using Randomization to Solve the Selection Problem | 70
- 3.6 Introducing EPATE: The Case When $\tau_{P_i=1} \neq \tau$ | 72
- 3.7 Recovering and Interpreting Heterogeneity of Treatment Effects | 74
- 3.8 Violations of the Exclusion Restrictions | 77
 - 3.8.1 SUTVA | 77
 - 3.8.2 Observability | 79
 - 3.8.3 Compliance | 80
 - 3.8.4 Statistical Independence | 81
- 3.9 Conclusions | 82

Appendix 3.1: Recovering the Wedge between τ and $\tilde{\tau}$ (Derivation of Equation 3.5) | 83

Appendix 3.2: The Brass Tacks of Estimating the Effects of Training Programs | 85

Notes | 87

References | 87

4. Statistical Conclusion Validity: Measurement in Economic Experiments | 89

Key Ideas | 89

- 4.1 Two Running Examples | 90
- 4.2 Perspectives on Sampling Frameworks | 92

4.2.1	Subpopulations in the Superpopulation Framework	94
4.3	Estimating Treatment Effects and Making Inference	95
4.3.1	Motivating the Difference-in-Means Estimator for ATE Parameters	95
4.3.2	Single Hypothesis Testing and Statistical Power	97
4.3.3	Multiple Hypothesis Testing	98
4.3.3.1	Family-Wise Error Rate	99
4.3.3.2	Approaches to Controlling the FWER	101
	<i>Bonferroni Correction</i>	102
	<i>Holm Stepdown Correction</i>	104
	<i>List et al. FWER Correction</i>	104
4.3.4	Introducing the Difference-in-Differences Estimator for ATE Parameters	107
4.3.5	Introducing an Alternative to ATE Parameters: Fisher's Randomization Inference	109
4.4	Conclusions	111
	Appendix 4.1: Code for List, Shaikh, and Xu (2019) and List, Shaikh, and Vayalinkal (2023)	112
	Notes	115
	References	116

PART II. DESIGNING ECONOMIC EXPERIMENTS

5. Optimal Experimental Design | 121

Key Ideas | 121

5.1	Three Running Examples	122
5.2	Basic Principles of Statistical Power	124
5.3	The Case of a Binary Treatment with Continuous Outcomes	125
5.3.1	Putting It All Together to Create an Optimal Design	127
5.4	The Case of a Binary Treatment with Binary Outcomes	131
5.5	Varying Treatment Levels with Continuous Outcomes	132
5.6	Expanding the Tool Kit	135
5.6.1	Heterogeneity in Participant Costs	135
5.6.2	Clustered Experimental Designs	137
5.6.3	Optimal Design with Multiple Hypothesis Adjustment	141
5.7	Less Considered Design Choices to Enhance Statistical Power	142
5.7.1	Including Covariates in the Estimation Model	143
5.7.2	Designs to Maximize Compliance	145
5.7.3	The Nature of the Sample	147
5.7.4	Measurement Choices	148
5.7.5	Factorial Designs	148
5.8	Conclusions	149

Appendix 5.1: An Example of the Power of Simulation Methods: The Case of Varying Treatment Levels with Binary Outcomes	150
Appendix 5.2: Step-by-Step Flexible Regression Adjustment	152
Appendix 5.3: Introducing Full and Fractional Factorial Designs	159
Appendix 5.4: A Walk-Through Example	166
Notes	168
References	170
6. Randomization Techniques	 173
Key Ideas	173
6.1 Three Running Examples	174
6.2 Classical Assignment Mechanisms	176
6.3 Classical Randomization Approaches	177
6.3.1 Bernoulli Trials	178
6.3.2 Completely Randomized Experiments (CRE)	180
6.3.3 Randomized Block (Stratified) Experiments	186
6.3.4 Rerandomization Approaches	189
6.3.5 Optimal Stratification with Matched-Pairs Designs	192
6.3.5.1 Efficient Matching Minimizing Mean-Squared Error	192
6.4 Design-Conscious Inference	195
6.4.1 Statistical Inference in CREs	195
6.4.2 Adjusting Inference under Alternative Randomization Schemes	196
6.5 What to Do with Unanticipated Covariates	197
6.6 Conclusions	197
Appendix 6.1: A Review of Rerandomization Approaches	198
Notes	201
References	202
7. Heterogeneity and Causal Moderation	 205
Key Ideas	205
7.1 Four Running Examples	207
7.2 Estimating Heterogeneities in Simple Cases	209
7.2.1 Using Causal Forests to Estimate Heterogeneities	212
Eight-Step Causal Forest Procedure	214
7.3 Basic Mechanics of Causal Moderation	216
7.3.1 Causal Moderation in Economic Experiments	219
7.4 Two Crucial Margins of Heterogeneity: Intensive and Extensive	222
7.4.1 Bounding the Intensive and Extensive Margin Effects	224
7.4.2 Using Baseline Outcome Data to Identify Intensive Margin Effects	226
7.4.3 A Tobit Approach to Estimating Margins	227
7.5 Conclusions	229

Notes | 229

References | 230

8. Mediation: Exploring Relevant Mechanisms | 233

Key Ideas | 233

8.1 Three Running Examples | 234

8.2 Mediation: The Basics of Causal Pathways | 237

8.2.1 Decomposing Total Effects in the Presence of Mediators | 240

8.2.2 Moving the Goalposts: Controlled and Principal-Strata Effects | 241

8.3 Applied Mediation Analysis for Economic Experiments | 244

8.3.1 A Parametric Workhorse and Its Pitfalls | 244

8.3.2 Basic Case: Binary Randomized Treatment | 245

8.3.3 Separate Randomization of Treatment and Mediator | 246

8.3.4 Paired Design | 248

8.3.5 Crossover Design | 249

8.4 Conclusions | 251

Appendix 8.1: Putting It All Together: Traditional Mediation Analysis and Alternative Approaches Using an In-Home Parent Visitation Program | 252

Notes | 257

References | 258

9. Experiments with Longitudinal Elements | 261

Key Ideas | 261

9.1 Three Running Examples | 263

9.2 Potential Outcomes in Repeated Exposure Designs | 264

9.2.1 Treatment Effects in the Presence of Repeated Exposures | 265

9.3 Staggered Experimental Design | 266

9.4 Leveraging Pre- and Post-treatment Outcomes to Increase Power | 267

9.4.1 Including Covariates and Pre-treatment Outcomes in the Estimation Model | 268

9.4.2 Leveraging Pre-treatment Outcomes in a Panel Data Estimation Model | 268

9.4.2.1 Gains from Pre-treatment Outcome Measures | 269

9.4.2.2 Autocorrelations That Vary with Treatment | 270

9.4.3 Choosing the Optimal Number of Pre-treatment and Post-treatment Periods | 272

9.4.4 Threats to Internal Validity | 272

9.5 Experimental Designs with Outcomes Measured Long after Treatment | 274

9.5.1 Identification Assumptions When Outcomes Are Far Removed from Treatment | 277

9.5.2	Statistical Surrogates 278
9.5.2.1	Internal Validity of Statistical Surrogates 279
9.5.2.2	Putting the Comparability and Surrogacy Assumptions into Perspective 281
9.5.2.3	Interpreting Surrogates 282
9.5.2.4	Multiple Surrogates 282
9.6	Conclusions 285
Appendix 9.1: Optimal Staggered Designs 286	
Appendix 9.2: Clustered Design in Panel Data Settings 288	
Appendix 9.3: Cluster-Randomized Experiments in Settings That Generate Short Panel Data 292	
Notes 292	
References 293	

10. Within-Subject Experimental Designs | 297

Key Ideas 297	
10.1	Three Running Examples 299
10.2	Potential Outcomes in a Within-Subject Design 300
10.3	Identification Assumptions in a Within-Subject Design 305
10.4	Threats to the Internal Validity of Within-Subject Designs 307
10.4.1	Threats to Balanced Panel 307
10.4.2	Threats to Temporal Stability 308
10.4.2.1	Crossover Designs and Latin Squares 310
10.4.3	Threats to Causal Transience 311
10.4.3.1	Washout Periods 312
10.5	Key Advantages of Within-Subject Designs 313
10.5.1	Heterogeneity and the Full Distribution of Treatment Effects 313
10.5.2	Experimental Power 315
10.5.2.1	Minimum Detectable Effects for Within-Subject Designs 316
10.6	Conclusions 318
Notes 319	
References 320	

PART III. VIOLATIONS OF EXCLUSION RESTRICTIONS

11. SUTVA: Interference and Hidden Treatments | 325

Key Ideas 325	
11.1	Three Running Examples 326
11.2	SUTVA Violation: Interference 328

11.2.1	Treatment Effect Parameters	332
11.2.2	Difference-in-Means	334
11.3	Approaches to Dealing with Interference Violations	334
11.3.1	Linear-in-Means Model	334
11.3.2	Clustered Randomized Trials to Attenuate Spillovers	336
11.3.3	Randomization Inference under Interference	337
11.4	Embracing Spillovers: Randomized Saturation Designs	340
11.4.1	Designs to Explore Spillovers	340
11.5	Hidden Versions of Treatment	343
11.5.1	Potential Outcomes with Hidden Versions of Treatment	344
11.5.2	Implications of Hidden Versions of Treatment	345
11.6	Conclusions	345
	Appendix 11.1: Optimal Saturation Design	346
	Notes	349
	References	349

12. Observability: Nonrandom Attrition | 351

Key Ideas | 351

12.1	Two Running Examples	352
12.2	Attrition in the Potential Outcomes Framework	354
12.2.1	Internal Validity for Respondents	355
12.2.2	Internal Validity for Study Participants	356
12.3	Tests for Internal Validity	358
12.3.1	Tests Using Baseline Outcome Data	358
12.3.2	Selective Attrition Test	359
12.3.3	Determinants of Attrition Test	361
12.3.4	Attrition Rates That Vary by Treatment	362
12.4	Analyzing Data with Attrition	365
12.4.1	Available Case Analysis	366
12.4.2	Horowitz and Manski Bounds	369
12.4.3	Inverse Probability Weighting	370
12.4.4	Selection Models	372
12.4.5	Lee Bounds	375
12.5	Missing Covariates	378
12.5.1	Complete and Available Case Analysis	378
12.5.2	Dummy Variable Adjustment	378
12.5.3	Imputation	379
12.6	Six Design Tips to Attenuate Attrition	380
12.7	Conclusions	382

Appendix 12.1: Putting It All Together with CHECC | 383

Notes | 402

References | 403

13. Complete Compliance: One-Sided and Two-Sided Violations | 407

Key Ideas | 407

13.1 Two Running Examples | 408

13.2 A Framework for Imperfect Compliance | 410

13.2.1 As-Treated Analysis Reintroduces the Selection Problem | 411

13.2.2 Intention-to-Treat (ITT) Analysis | 413

13.3 Randomization as an Instrumental Variable and New Assumptions | 415

13.4 Calculating ATEs for Compliers | 418

13.4.1 Characterizing Compliers | 422

13.4.2 Widening the Goalposts: Bounding the ATE | 424

13.5 Six Design Tips to Attenuate Noncompliance | 426

13.6 Conclusions | 428

Appendix 13.1: Encouragement Designs | 430

Notes | 434

References | 436

14. Statistical Independence and Compromised Randomization | 439

Key Ideas | 439

14.1 Three Running Examples | 441

14.2 Statistical Independence: The Basics | 444

14.3 Tests for Compromised Randomization | 445

14.3.1 Comparing Planned versus Actual Assignment | 445

14.3.2 Computing P-Values to Test for Compromised
Randomization | 447

14.3.3 Informal Checks of Compromised Randomization | 449

14.4 Case 1: A Rerandomization Approach | 449

14.5 Case 2a: Inference with Compromised Randomization and Full
Documentation | 45114.5.1 Inference When the Randomization Procedure Is Correlated with
Potential Outcomes | 45214.6 Case 2b: Inference with Compromised Randomization and Only Partial
Documentation | 45614.6.1 An Example of Compromised Randomization Being Partly
Understood at the Aggregate Level | 456

14.6.2 Breaking Down the Randomization Procedure | 457

14.6.3 A Basic Model | 458

14.6.4 Testing a Single Joint Null Hypothesis | 461

14.7 A Decision-Theoretic Framework with Incomplete Documentation | 463

14.7.1 Modeling the Randomization Protocol | 464

14.7.2 Partially Identifying Model Parameters | 465

14.7.3 Worst-Case Randomization Test | 466

- 14.8 Seven Design Tips to Prevent Compromised Randomization | 468
 - 14.8.1 Three Tips When the Researcher Is Responsible for Randomization | 468
 - 14.8.2 Four Tips When the Experimenter Relies on Partners for Randomization | 469
- 14.9 Conclusions | 470
- Appendix 14.1: Using Fisher's *Sharp* Inference with Compromised Randomization | 470
- Appendix 14.2: Putting the Ideas of Section 14.6 in Motion | 472
- Appendix 14.3: Extending Section 14.6 to Test Multiple Hypotheses | 473
- Notes | 474
- References | 476

PART IV. BUILDING SCIENTIFIC KNOWLEDGE

15. Building Confidence in (and Knowledge from) Experimental Results | 481

Key Ideas | 481

- 15.1 Three Running Examples | 483
- 15.2 The Philosophy of Building Knowledge from Experimental Results | 485
- 15.3 A Framework for Building Confidence in Experimental Results | 487
 - 15.3.1 Effects of α and β on the PSP | 490
 - 15.3.2 Null Results Are Informative Too | 493
- 15.4 From the Researcher to the Research Community | 497
 - 15.4.1 Replication Types | 497
 - 15.4.1.1 Interpreting Replication Results | 500
 - 15.4.1.2 Building Knowledge and Confidence with Replications | 500
 - 15.4.1.3 Why Are Replications an Endangered Species in Economics? | 503
- 15.5 The Beauty of Selective Data Generation: From the Lab to the Field | 505
- 15.6 Conclusions | 509
- Appendix 15.1: Gaining Insights into Equation 15.5 and Beyond | 510
 - Unbiased, Sympathetic, and Adversarial Replications | 511
 - Heterogeneity across Replicating Teams | 512
 - Should We Have Confidence in Our Updating from Experimental Results? | 513
- Notes | 515
- References | 517

16. Generalizability and Scaling | 523

Key Ideas | 523

16.1 Two Running Examples | 525

16.2 External Validity Primers | 526

16.2.1 From Treatment Effects to the Parameter of Interest | 527

16.2.2 Three Types of Horizontal Generalizability | 529

16.2.3 Assumptions Yielding $\tau = \tau^*$ | 530

16.3 Digging Deeper into Assumptions 16.1-16.4 | 533

16.3.1 Assumption 16.1: Selection into the Experiment | 534

16.3.1.1 A Model of Selection into Experiments | 534

16.3.1.2 How Experimental Design Affects Selection | 537

16.3.2 Assumption 16.2: Representativeness of the Population | 539

16.3.3 Assumptions 16.3 and 16.4: Investigation Neutrality and
Parallelism | 54016.3.3.1 Experimenter Scrutiny: Effects of A | 54216.3.3.2 Experimental Environment: Effects of E | 54316.3.3.3 Stakes: Effects of I_i | 545

16.4 Scaling | 545

16.4.1 A Behavioral Model of Scaling | 546

16.4.2 Constructive Steps Forward: The SANS Conditions | 549

Author Onus Probandi | 549

16.4.3 Three Waves of Scientific Research | 550

16.5 Conclusions | 553

Appendix 16.1: Mechanics of Scaling Up | 553

Notes | 556

References | 558

PART V. THE ETHICAL AND PRACTICAL SIDES OF ECONOMIC EXPERIMENTS**17. The Ethics of Economic Experiments | 565**

Key Ideas | 565

17.1 Four Running Examples | 567

17.2 Ethics Primer | 570

17.2.1 A Simple Economic Model | 570

17.2.2 A Simple Philosophical Framework | 572

17.3 Three Theories of (Research) Ethics | 576

17.3.1 Consequentialism | 576

17.3.2 Deontological Ethics | 577

17.3.3 Rule Consequentialism | 578

17.4	Putting It All Together 579
17.4.1	Truthful, Unbiased, and Transparent Reporting of Results and Conflicts of Interest 579
17.4.2	Appropriate Data Governance and Management 581
17.4.3	Conflicts between Individual Protections and Scientific Discovery 583
17.4.3.1	<i>Should You Even Do an Experiment?</i> 584
17.4.3.2	<i>With Whom Should You Experiment?</i> 587
17.4.3.3	<i>How Should You Experiment?</i> 588
17.4.3.3.1	<i>Informed Consent: Respecting Autonomy</i> 588
17.4.3.3.2	<i>Defining Benefits and Harm: From the Subject to Innocent Bystanders</i> 591
17.4.3.3.3	<i>Outright Deception and Incomplete Disclosure</i> 594
17.5	Benchmarking Research Ethics: Gold to Plutonium-239 597
17.6	Conclusions 603
	Appendix 17.1: Data Governance and Management Playbook 603
	Being Trustworthy for Knowledge Creation 604
	Being Trustworthy regarding Subjects 604
	Differential Privacy 605
	Being Trustworthy for Third Parties 605
	Accessibility and Accountability 605
	Security 606
	Notes 606
	References 609
	18. Pre-treatment Administrative Responsibilities 617
	Key Ideas 617
18.1	One Running Example 619
18.2	Overarching Goals of Pre-treatment Tasks 621
18.3	Institutional Review 622
18.3.1	IRBs and Research Ethics 622
18.3.2	IRB Application Materials 623
18.3.2.1	IRB Requirements: Who, What, How, and to Whom? 624
18.3.3	IRB Review Process and Determinations 625
18.3.3.1	IRBs and Informed Consent 627
18.3.3.2	IRBs and Outright Deception 628
18.3.3.3	IRBs and Pilots 629
18.3.3.4	IRBs and Multi-institutional Research 630
18.3.3.5	Communication with IRBs 630

18.4	Registries and Pre-analysis Plans	632
18.4.1	Trial Registries	633
18.4.1.1	Existing Registries	634
18.4.1.2	The AEA Registry	635
18.4.1.3	Registry Limitations	636
18.4.2	Pre-analysis Plans	636
18.5	Data Use Agreements and Outside Partners	642
18.5.1	Components of a DUA	643
18.6	Due Diligence Administrative Checklist	644
18.7	Conclusions	644
	Appendix 18.1: A Plea to the IRB	645
	Notes	647
	References	648

19. Optimal Use of Incentives in Economic Experiments | 651

Key Ideas | 651

19.1	Four Running Examples	653
19.2	A Simple Economic Model	655
19.2.1	Extending the Model to Explore Knowledge Creation: Internal Validity	657
19.2.1.1	Within-Subject versus Between-Subject Design	660
19.2.1.2	Statistical Surrogates	662
19.2.2	Extending the Model to Explore Knowledge Creation: Improving Inference	663
19.2.2.1	Nuts and Bolts of Design	663
19.2.2.2	Pilot Experiments	664
19.2.2.3	Mediators and Moderators	665
19.2.2.4	EP2: From $\tau_{P_i=1}$ to τ and Beyond	667
19.2.2.5	EP2: From One Environment to Another	670
19.2.2.6	EP2: Fostering Scaling by Adding Option C Thinking to Designs	670
19.3	Creating the Microeconomic Environment	672
19.3.1	Using Induced Values for Control	673
19.3.2	Potentially Losing Control	675
19.3.2.1	An Inferential Challenge: Flat Payoffs	675
19.3.2.2	An Inferential Challenge: Construct Validity	680
19.3.2.3	Experimental Instructions across the Empirical Spectrum	680
19.4	Conclusions	682
	Appendix 19.1: Inducing Risk Posture	683
	Appendix 19.2: Tips for Writing Experimental Instructions across the Empirical Spectrum	684
	10 Tips for Writing Laboratory Experimental Instructions	684

From the Lab to the Field 685
8 Tips for Artefactual Field Experiments (AFEs) 685
6 Tips for Framed Field Experiments (FFE) 687
5 Tips for Natural Field Experiments (NFE) 687
Practical Implementation 689
Conclusion 689

Notes 689
References 691

20. Epilogue: The (Written) Road to Scientific Knowledge Diffusion | 695

Key Ideas 695

20.1 Give the People What They Want! But . . . What Do They Want? 697
20.2 Creating a Logical Framework 698
20.2.1 Applying BEC Holistically 699
PREP 699
20.3 Your Writing Style 706
20.3.1 Getting Started: An Eight-Step “Inside-Out Approach” to Writing Scientific Studies 708
20.4 Introducing Your Pen to the World 710
20.5 Epilogue 712

Appendix 20.1: PREP Checklist: Proper Reporting in an Experimental Paper 713
Notes 714
References 715

PART VI. “HOW-TO” SUPPLEMENTS

S1. How to Conduct Experiments in Markets: From the Lab to the Field 717
S2. How to Conduct Experiments with Organizational Partnerships 717
S3. How to Conduct Experiments with Children 717
S4. How to Conduct Experiments to Measure Preferences, Beliefs, and Constraints 717
Glossary 719
Notation Crib Sheet 729
Further Readings 741
Index 847