

Matching, Regression Discontinuity, Difference in Differences, and Beyond

Myoung-jae Lee,

OXFORD
UNIVERSITY PRESS

CONTENTS

Preface	xv
1 Basics of Treatment Effect Analysis	1
1.1 Counterfactual, Intervention, and Causal Relation	1
1.1.1 Potential Outcomes and Intervention	1
1.1.2 Causality and Association	3
1.1.3 Partial Equilibrium Analysis and Remarks	4
1.2 Various Treatment Effects and No Effects	5
1.2.1 Various Effects	5
1.2.2 Three No-Effect Concepts	6
1.2.3 Remarks	7
1.3 Group-Mean Difference and Randomization	8
1.3.1 Group-Mean Difference and Mean Effect	8
1.3.2 Consequences of Randomization	10
1.3.3 Checking Out Covariate Balance	12
1.4 Overt Bias, Hidden Bias, and Selection Problems	13
1.4.1 Overt and Hidden Biases	13
1.4.2 Selection on Observables and Unobservables	14
1.4.3 Linear Models and Biases	16
1.5 Estimation with Group Mean Difference and LSE	18
1.5.1 Group-Mean Difference and LSE	18
1.5.2 Job Training Example	19
1.5.3 Linking Counterfactuals to Linear Models	21
1.6 Structural Form, Assignment, and Marginal Model	22
1.6.1 Structural versus Reduced Forms for Response	22
1.6.2 Treatment Structural Form and Assignment	23
1.6.3 Marginal Structural Model	24
1.7 Simpson's Paradox and False Covariate Control	25
2 Matching	28
2.1 Basics of Matching and Various Effects	28
2.1.1 Main Idea	28
2.1.2 Effect on Treated and Effect on Population	30
2.1.3 Dimension and Support Problems	31
2.1.4 Variables to Control	32

2.2	Implementing Matching	35
2.2.1	Decisions to Make in Matching	35
2.2.2	Matching Estimators	37
2.2.3	Asymptotic Variance Estimation	40
2.2.4	Labor Union Effect on Wage	44
2.3	Propensity Score Matching (PSM)	46
2.3.1	Propensity Score as a Balancing Score	46
2.3.2	Removing Overt Bias with Propensity Score	47
2.3.3	Implementing PSM and Bootstrap	48
2.3.4	PSM Empirical Examples	50
2.3.5	Propensity Score Specification Issues*	52
2.4	Further Remarks	54
2.4.1	Covariate Balance Check	54
2.4.2	Matching for Hidden Bias	56
2.4.3	Prognostic Score and More*	58
3	Nonmatching and Sample Selection	61
3.1	Weighting	61
3.1.1	Weighting Estimator for Effect on Population	61
3.1.2	Other Weighting Estimators and Remarks	63
3.1.3	Asymptotic Distribution of Weighting Estimators*	65
3.1.4	Job Training Effect on Unemployment	66
3.1.5	Doubly Robust Estimator*	67
3.1.6	Weighting for Missing Data*	68
3.2	Regression Imputation	69
3.2.1	Linear Regression Imputation	70
3.2.2	Regression Imputation with Propensity Score	71
3.2.3	Regression Imputation for Multiple Treatment	72
3.2.4	Regression Imputation for Continuous Treatment*	73
3.2.5	Military Service Effect on Wage	74
3.3	Complete Pairing with Double Sum	76
3.3.1	Discrete Covariates	77
3.3.2	Continuous Covariates	79
3.3.3	Nonparametric Distributional Effect Tests*	80
3.4	Treatment Effects under Sample Selection	84
3.4.1	Difficulties with Sample Selection Models	85
3.4.2	Participation, Invisible, and Visible Effects	86
3.4.3	Identification of Three Effects with Mean Differences	87
3.4.4	Religiosity Effect on Affairs	88
3.5	Effect Decomposition in Sample Selection Models*	90
3.5.1	Motivation for Decomposition	90
3.5.2	Decomposition with Linear Selection Model	91
3.5.3	Four Special Models	92
3.5.4	Race Effect on Wage	94

4	Regression Discontinuity	97
4.1	Introducing RD with Before-After	97
4.1.1	BA Examples	97
4.1.2	BA Identification Assumption	98
4.1.3	From BA to RD	99
4.2	RD Identification and Features	100
4.2.1	Sharp RD (SRD) and Fuzzy RD (FRD)	101
4.2.2	Identification at Cutoff	102
4.2.3	RD Main Features	104
4.2.4	Class Size Effect on Test Score	106
4.3	RD Estimators	109
4.3.1	LSE for Level Equation	109
4.3.2	IVE for Right-Left Differenced Equation	110
4.3.3	Bandwidth Choice and Remarks	112
4.3.4	High School Completion Effect on Fertility	113
4.4	Specification Tests	116
4.4.1	Breaks in Conditional Means	116
4.4.2	Continuity in Score Density	117
4.5	RD Topics*	119
4.5.1	Spatial Breaks	119
4.5.2	RD for Limited Dependent Variables	120
4.5.3	Measurement Error in Score	121
4.5.4	Regression Kink (RK) and Generalization	123
4.5.5	SRD with Multiple Scores	126
4.5.6	Quantile RD	129
5	Difference in Differences	131
5.1	DD Basics	131
5.1.1	Examples for DD	132
5.1.2	Time-Constant and Time-Varying Qualifications	133
5.1.3	Data Requirement and Notation	135
5.2	DD with Repeated Cross-Sections	136
5.2.1	Identification	136
5.2.2	Identification with Parametric Models	140
5.2.3	Schooling Effect on Fertility: 'Fuzzy DD'	142
5.2.4	Linear Model Estimation for Two Periods or More	144
5.2.5	Earned Income Tax Credit Effect on Work	147
5.2.6	Time-Varying Qualification*	148
5.3	DD with Panel Data	150
5.3.1	Identification	150
5.3.2	Identification and Estimation with Parametric Models	152
5.3.3	Daylight Saving Time Effect on Energy	157
5.4	Panel Stayer DD for Time-Varying Qualification	158
5.4.1	Motivation	158

5.4.2	Effect on In-Stayers Identified by Stayer DD	159
5.4.3	Identification and Estimation with Panel Linear Models	160
5.4.4	Pension Effect on Health Expenditure	162
6	Triple Difference and Beyond	165
6.1	TD Basics and More	165
6.2	TD with Repeated Cross-Sections	166
6.2.1	Identification	166
6.2.2	Identification and Estimation with Linear Models	169
6.2.3	Mandated Benefit Effect on Wage	172
6.3	TD with Panel Data	174
6.3.1	Identification	174
6.3.2	Estimation with Panel Linear Model	175
6.3.3	Tax-Inclusive Price Effect on Demand	177
6.4	GDD and Beyond	178
6.4.1	Motivation for GDD and Beyond	179
6.4.2	Identification for GDD and QD	180
6.4.3	Identified Effects When Panel Linear Model Holds	181
6.4.4	LSE for DD and GDD and Testing for DD Condition	182
6.4.5	Sulfa Drug Effect on Mortality: Is DD Trustworthy?	184
6.5	Clustering Problems and Inference for DD and TD	187
6.5.1	Single Clustering	188
6.5.2	Clustering in Panel Data	194
6.5.3	DD and TD with Cluster-Specific Treatment	199
6.5.4	Details on Cluster Variance Estimator*	202
A	APPENDIX	209
A.1	Kernel Density and Regression Estimators	209
A.1.1	Histogram-Type Density Estimator	209
A.1.2	Kernel Density Estimator	210
A.1.3	Kernel Regression Estimator	211
A.1.4	Local Linear Regression	213
A.2	Bootstrap	213
A.2.1	Review on Usual Asymptotic Inference	215
A.2.2	Bootstrap to Find Quantiles	216
A.2.3	Percentile-t and Percentile Methods	218
A.2.4	Nonparametric, Parametric, and Wild Bootstraps	219
A.3	Confounder Detection, IVE, and Selection Correction	220
A.3.1	Coherence Checks	220
A.3.2	IVE and Complier Effect	225
A.3.3	Selection Correction Approach	230
A.4	Supplements for DD Chapter	232
A.4.1	Nonparametric Estimators for Repeated Cross-Section DD	233
A.4.2	Nonparametric Estimation for DD with Two-Wave Panel Data	233

A.4.3	Panel Linear Model Estimation for DD with One-Shot Treatment	236
A.4.4	Change in Changes	238
	References	241
	Index	253
Online GAUSS Programs:		
	Pair Matching with PS for Union on Wage (PairMatchUnionOnWage)	
	Regression Imputation with PS-Based Nonparametrics (RegImpPsNprSim)	
	Complete Pairing with PS for Union on Wage (CpUnionOnWage)	
	RD Program (RdSim)	
	Repeated Cross-Section DD (DdReCroVary4WavesSim)	
	Panel DD for Differenced Model (DdPanel6WavesSim)	
	Repeated Cross-Section TD (TdReCro2WavesSim)	
	Panel DD and GDD for Differenced Model (DdGddPanel5WavesSim)	
	Panel DD, GDD and QD for Sulfa Drug (DdGddQdSulfaDrug)	
	Bootstrap for Sample Mean (BootAvgSim)	
	Selection Correction for Work on Doctor Visits (SelCorcWorkOnVisit)	
	Panel LSE, WIT and BET (PanelLseWitBetSim)	