Full text available at: http://dx.doi.org/10.1561/0800000014

The Estimation of Causal Effects by Difference-in-Difference Methods

Full text available at: http://dx.doi.org/10.1561/0800000014

The Estimation of Causal Effects by Difference-in-Difference Methods

Michael Lechner

Swiss Institute for Empirical Economic Research
University of St. Gallen
Switzerland
Michael.Lechner@unisg.ch

the essence of knowledge

Boston – Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is M. Lechner, The Estimation of Causal Effects by Difference-in-Difference Methods, Foundations and Trends® in Econometrics, vol 4, no 3, pp 165–224, 2010

ISBN: 978-1-60198-498-2

Foundations and Trends® in Econometrics

Volume 4 Issue 3, 2010

Editorial Board

Editor-in-Chief:
William H. Greene
Department of Economics
New York Univeristy
44 West Fourth Street, 7–78
New York, NY 10012
USA
wgreene@stern.nyu.edu

Editors

Editorial Scope

Foundations and Trends® in Econometrics will publish survey and tutorial articles in the following topics:

- Identification
- Model Choice and Specification Analysis
- Non-linear Regression Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Dynamic Specification
- Inference and Causality
- Continuous Time Stochastic Models
- Modeling Non-linear Time Series
- Unit Roots
- Cointegration
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Interactions-based Models
- Duration Models
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians
Foundations and Trends® in Econometrics, 2010, Volume 4, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Foundations and Trends® in
Econometrics
Vol. 4, No. 3 (2010) 165–224

DOI: 10.1561/0800000014

now
the essence of knowledge

The Estimation of Causal Effects by Difference-in-Difference Methods*

Michael Lechner

Professor of Econometrics, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstrasse 14, CH-9000 St. Gallen, Switzerland, Michael.Lechner@unisg.ch

Abstract

This survey gives a brief overview of the literature on the difference-in-difference (DiD) estimation strategy and discusses major issues using a treatment effects perspective. In this sense, this survey gives a somewhat different view on DiD than the standard textbook discussion of the DiD model, but it will not be as complete as the latter. It contains some extensions of the literature, for example, a discussion of, and suggestions for nonlinear DiD estimators as well as DiD estimators based on propensity-score type matching methods.

* I am a Research Fellow of CEPR, London, CESifo, Munich, IAB, Nuremberg, IZA, Bonn, and PSI, London. This project received financial support from the St. Gallen Research Centre of Aging, Welfare, and Labour Market Analysis (SCALA). The paper was presented at the econometrics breakfast at the University of St. Gallen and benefited from comments by seminar participants. I thank Christina Felfe, Alex Lefter, Conny Wunsch as well as Hans Fricke and Susanne Zopf for carefully reading a previous draft of this manuscript, Patrick Puhani for an interesting discussion on some issues related to limited dependent outcome variables, and Jeff Smith for additional useful references. The comments of a referee of this journal helped to improve the paper in several dimensions. The usual disclaimer applies.

Keywords: Causal inference, counterfactual analysis, before-after-treatment-control design, control group design with pretest and posttest.

JEL Codes: C21, C23, C31, C33

Contents

1

Introduction

The Difference-in-Difference (DiD) approach is a research design for estimating causal effects. It is popular in empirical economics, for example, to estimate the effects of certain policy interventions and policy changes that do not affect everybody at the same time and in the same way. It is used in other social sciences as well.[1] DiD could be an attractive choice when using research designs based on controlling for confounding variables or using instrumental variables is deemed unsuitable, and at the same time, pre-treatment information is available.[2] The DiD design is usually based on comparing *de facto* four different groups of objects. Three of these groups are not affected by the treatment. In many applications, "time" is an important variable to

[1] In other social sciences the DiD approach is also denoted as "untreated control group design with independent pretest and posttest samples" or "control group design with pretest and posttest." See, for example, Cook and Campbell (1979), Rosenbaum (2001), and Shadish et al. (2002) for further references.

[2] Following the literature, the event for which we want to estimate the causal effect is called the *treatment*. The *outcome* denotes the variable that will be used to measure the effect of the treatment. Outcomes that would be realised if a specific treatment has, or would have been applied, are called *potential outcomes*. A variable is called *confounding* if it is related to the treatment and the potential outcomes. A variable is called an *instrument* if it influences the treatment but not the potential outcomes.

distinguish the groups.[3] Besides the group which already received the treatment (post-treatment treated), these groups are the treated prior to their treatment (pre-treatment treated), the nontreated in the period before the treatment occurs to the treated (pre-treatment nontreated), and the nontreated in the current period (post-treatment nontreated).[4] The idea of this empirical strategy is that if the two treated and the two nontreated groups are subject to the same time trends, and if the treatment has had no effect in the pre-treatment period, then an estimate of the "effect" of the treatment in a period in which it is known to have none, can be used to remove the effect of confounding factors to which a comparison of post-treatment outcomes of treated and nontreated may be subject to. This is to say that we use the mean changes of the outcome variables for the nontreated over time and add them to the mean level of the outcome variable for the treated prior to treatment to obtain the mean outcome the treated would have experienced if they had not been subjected to the treatment.

This survey presents a brief overview of the literature on the difference-in-difference estimation strategy and discusses major issues mainly using a treatment effect perspective (and language) that allows, in our opinion, more general considerations than the classical regression formulation that still dominates the applied work. In this sense, this survey might give a somewhat different perspective than the standard text book discussion of the DiD design, but it will not be as complete as the latter. Thus, this review is more of a complement than a substitute to the excellent text type discussions of the DiD approach that are already available in the literature (e.g., Angrist and Pischke, 2009; Blundell and Costa Dias, 2009; Imbens and Wooldridge, 2009).

This review focuses on the case of only two differences although the basic ideas of DiD estimation could be extended to more than

[3] As the concept of time is only used to define a group that is similar to the treated group with respect to relevant unobservable variables and whose members have not (yet) participated, any other characteristic may be used instead of time as well, as long as the formal conditions given below are fulfilled.

[4] When a data set is available in which everybody is observed in all periods, there will be just two groups with outcomes measured before and after the treatment.

two dimensions to create difference-in-difference-in-difference-in- ... estimators.[5] However, the basic ideas of the approach of taking multiple differences are already apparent with two dimensions. Thus, we refrain from addressing these higher dimensions to keep the discussion as focused as possible.

The outline of this survey is as follows: The next section gives a historical perspective and discusses some interesting applications. Section 3, which is the main part of this survey, discusses identification issues at length. Section 4 concerns DiD specific issues related to estimation, including a discussion of propensity score matching estimation of DiD models. Section 5 discusses some specific issues related to inference, and Section 6 considers important practical extensions to the basic approach. Section 7 concludes. Some short proofs are relegated to a technical appendix.

[5] For example, Yelowitz (1995) analyzes the effects of losing public health insurance on labor market decisions in the United States by using Medicaid eligibility that varies over time, state and age (of the child in the household). Another example for a triple difference is the paper by Ravallion et al. (2005) who analyze the effects of a social programme based on a comparison of participants with nonparticipants and ex-participants.

References

Abadie, A. (2005), 'Semiparametric difference-in-difference estimators'. *Review of Economic Studies* **72**, 1–19.

Abadie, A., A. Diamond, and J. Hainmüller (2007), 'Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program'. NBER technical working paper 335.

Acemoglu, D. and J. D. Angrist (2001), 'Consequences of employment protection? The case of the Americans with disabilities act'. *Journal of Political Economy* **109**, 915–957.

Ai, D. and E. C. Norton (2003), 'Interaction terms in logit and probit models'. *Economics Letters* **80**, 123–129.

Angrist, J. D. and A. B. Krueger (1999), 'Empirical strategies in labor economics'. In: O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, vol. III A, Ch 23. pp. 1277–1366.

Angrist, J. D. and J.-S. Pischke (2009), *Mostly Harmless Econometrics*. New York: Princeton University Press.

Ashenfelter, O. (1978), 'Estimating the effect of training programs on earnings'. *The Review of Economics and Statistics* **60**(1), 47–57.

Ashenfelter, O. and D. Card (1985), 'Using the longitudinal structure of earnings to estimate the effect of training programs'. *The Review of Economics and Statistics* **67**, 648–660.

Athey, S. and G. W. Imbens (2006), 'Identification and inference in nonlinear difference-in-difference models'. *Econometrica* **74**, 431–497.

Autor, D. H. (2003), 'Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing'. *Journal of Labor Economics* **21**, 1–42.

Bergemann, A., B. Fitzenberger, and S. Speckesser (2009), 'Evaluating the dynamic employment effects of training programs in east Germany using conditional difference-in-differences'. *Journal of Applied Econometrics* **24**, 797–823.

Bertrand, M., E. Duflo, and S. Mullainathan (2004), 'How much should we trust differences-in-differences estimates'. *Quarterly Journal of Economics* pp. 249–275.

Besley, T. and R. Burgess (2004), 'Can labor regulation hinder economic performance? Evidence from India'. *Quarterly Journal of Economics* pp. 91–134.

Blundell, R. and M. Costa Dias (2009), 'Alternative approaches to evaluation in empirical microeconomics'. *Journal of Human Resources* **44**, 565–640.

Blundell, R., A. Duncan, and C. Meghir (1998), 'Estimating labor supply responses using tax reforms'. *Econometrica* **66**, 827–861.

Blundell, R., C. Meghir, M. Costa Dias, and J. van Reenen (2004), 'Evaluating the employment impact of a mandatory job search program'. *Journal of the European Economic Association* **2**, 569–606.

Bonhomme, S. and U. Sauder (2011), 'Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling'. *The Review of Economics and Statistics* **93**, 479–494.

Bühler, S., M. Helm, and M. Lechner (2011), 'Trade liberalization and growth: Plant-level evidence from Switzerland'. Discussion paper 2011-33, Department of Economics, University of St. Gallen.

Card, D. (1990), 'The impact of the mariel boatlift on the miami labor market'. *Industrial and Labor Relations Review* **43**(2), 245–257.

Card, D. and A. B. Krueger (1994), 'Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania'. *The American Economic Review* **84**, 772–793.

Chabé-Ferret, S. (2010), *To Control or Not to Control? Bias of Simple Matching vs Difference-In-Difference Matching in a Dynamic Framework*. mimeo.

Conley, T. and C. Taber (2011), 'Inference with "difference in differences" with a small number of policy changes'. *Review of Economics and Statistics* **93**, 113–125.

Cook, P. J. and G. Tauchen (1982), 'The effect of liquor taxes on heavy drinking'. *Bell Journal of Economics* **13**, 379–390.

Cook, T. D. and D. T. Campbell (1979), *Quasi-Experimentation*. Boston: Houghton Mifflin.

Donald, S. G. and K. Lang (2007), 'Inference with difference-in-differences and other panel data'. *Review of Economics and Statistics* **89**, 221–233.

Duflo, E. (2001), 'Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment'. *American Economic Review* **91**, 795–813.

Eichler, M. and M. Lechner (2002), 'An evaluation of public employment programmes in the east German state of Sachsen-Anhalt'. *Labour Economics: An International Journal* **9**, 143–186.

Eissa, N. (1996), 'Labor supply and the economic recovery act of 1981'. In: M. Feldstein and J. Poterba (eds.): *Empirical Foundations of Household Taxation*. pp. 5–38.

Gruber, J. and J. Poterba (1994), 'Tax incentives and the decision to purchase health insurance: Evidence from the self- employed'. *The Quarterly Journal of Economics* **109**, 701–733.

Hansen, C. B. (2007a), 'Asymptotic properties of a robust variance matrix estimator for panel data when T is large'. *Journal of Econometrics* **141**, 597–620.

Hansen, C. B. (2007b), 'Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects'. *Journal of Econometrics* **140**, 670–694.

Havnes, T. and M. Mogstad (2010), 'Is universal child care leveling the playing field? Evidence from a nonlinear difference-in-difference approach'. IZA discussion paper 4478.

Heckman, J. J. (1996), 'Comment on eissa: Labor supply and the economic recovery act of 1981'. In: M. Feldstein and J. Poterba (eds.): *Empirical Foundations of Household Taxation*. pp. 5–38.

Heckman, J. J. and V. J. Hotz (1989), 'Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training'. *Journal of the American Statistical Association* **84**, 862–880.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998), 'Characterizing selection bias using experimental data'. *Econometrica* **66**, 1017–1098.

Heckman, J. J., R. LaLonde, and J. Smith (1999), 'The economics and econometrics of active labor market programs'. In: O. Ashenfelter and D. Card (eds.): *Handbook of Labour Economics*, vol. 3. Amsterdam: North-Holland, pp. 1865–2097.

Heckman, J. J. and R. Robb (1986), 'Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes'. In: H. Wainer (ed.): *Drawing Inferences from Self-Selected Samples*. pp. 63–113.

Huber, M., M. Lechner, and C. Wunsch (2010), 'How to control for many covariates? Reliable estimators based on the propensity score'. Discussion paper 2010-30, Department of Economics, University of St. Gallen.

Hunt, J. (1995), 'The effect of unemployment compensation on unemployment duration in Germany'. *Journal of Labor Economics* **13**, 88–120.

Imbens, G. W. and J. M. Wooldridge (2009), 'Recent developments in the econometrics of program evaluation'. *Journal of Economic Literature* **47**, 5–86.

Lai, A. (2011), 'London cholera and the blind-spot of an epidemiology theory'. *Significance* pp. 82–85.

Lechner, M. (2008a), 'A note on endogenous control variables in evaluation studies'. *Statistics and Probability Letters* **78**, 190–195.

Lechner, M. (2008b), 'A note on the common support problem in applied evaluation studies'. *Annales d'Économie et de Statistique* **91–92**, 217–234.

Lester, R. A. (1946), 'Shortcomings of marginal analysis for the wage-employment problems'. *American Economic Review* **36**, 63–82.

Manski, C. F. (2011), 'Identification of treatment response with social interactions'. Department of Economics and Institute for Policy Research, Northwestern University, mimeo.

Meyer, B. D. (1995), 'Natural and quasi-experiments in economics'. *Journal of Business & Economic Statistics* **13**, 151–161.

Meyer, B. D., W. K. Viscusi, and D. L. Durbin (1995), 'Workers' compensation and injury duration: Evidence from a natural experiment'. *American Economic Review* **85**(3), 322–340.

Miguel, E. and M. Kremer (2004), 'Worms: Identifying impacts on education and health in the presence of treatment externalities'. *Econometrica* **72**, 159–217.

Obenauer, M. and B. von der Nienburg (1915), 'Effect of minimum-wage determinations in oregon'. *Bulletin of the U.S. Bureau of Labor Statistics*, 176, Washington, D.C.: U.S. Government Printing Office.

Puhani, P. (2008), 'The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-difference" models'. IZA Discussion paper 3478, revised November 2008.

Ravallion, M., E. Galasso, T. Lazo, and E. Philipp (2005), 'What can ex-participants reveal about a program's impact'. *Journal of Human Ressources* **40**, 208–230.

Rose, A. M. (1952), 'Needed research on the mediation of labour disputes'. *Personal Psychology* **5**, 187–200.

Rosenbaum, P. (2001), 'Stability in the absence of treatment'. *Journal of the American Statistical Association* **96**, 210–219.

Rosenbaum, P. R. and D. B. Rubin (1983), 'The central role of the propensity score in observational studies for causal effects'. *Biometrika* **70**, 41–50.

Rubin, D. B. (1977), 'Assignment to treatment group on the basis of a covariate'. *Journal of Educational Statistics* **2**, 1–26.

Shadish, W. R., T. D. Cook, and D. T. Campbell (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Simon, J. L. (1966), 'The price elasticity of liquor in the U.S. and a simple method of determination'. *Econometrica* **34**, 193–205.

Snow, J. (1854), 'The cholera near golden square, and at deptford'. *Medical Times and Gazette* **9**, 321–322.

Snow, J. (1855), *On the Mode of Communication of Cholera*. 2nd edition. London: John Churchill.

Waldfogel, J. (1998), 'The family gap for young women in the United States and Britain: Can maternity leave make a difference'. *Journal of Labor Economics* **16**, 505–545.

Yelowitz, A. S. (1995), 'The medicaid notch, labor supply, and welfare participation: Evidence from eligibility expansions'. *Quarterly Journal of Economics* **110**, 909–939.