

# Regional Research Institute West Virginia University

Resource Document Series



## Quasi-Experimental Methods an Annotated Bibliography

Mulugeta S. Kahsai and Randy Jackson

RRI Resource Doc 2015-01

Date submitted: June 22, 2015

Date revised:

Key words/Codes: Quasi experimental methods; Propensity Score  
Matching

# Quasi-Experimental Methods

Mulugeta S. Kahsai and Randy Jackson

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Literature</b>	<b>2</b>
<b>3</b>	<b>Alphabetical Listing of Sources and Abstracts</b>	<b>6</b>

## 1 Introduction

Quasi-experiments are defined as experiments that do not have random assignment but do involve manipulation of the independent variable. The key question addressed by this approach is “what would have happened had there been no policy/ treatment/ intervention”. To answer this fundamental question, it attempts to replicate traditional experimental design by comparing places in which a policy (treatment) has been implemented with carefully matched control groups in which no treatment has been implemented. The control group serves as the counterfactual for what would have happened to the treatment group in the absence of treatment. It provides a baseline forecast. This method is typically used in impact evaluations when experiments are infeasible.

Selecting the right control group and applying an appropriate impact evaluations method are key to the effectiveness of quasi-experimental approach. Over the years, researchers have developed different control group selection methods that use single/multiple similarities on the pre-treatment period. The most widely used are difference of means test, Mahalanobis distance, and propensity score. Impact evaluation is done based on simple difference of means test, simple regression, regression discontinuity, and difference of difference methods.

This annotated bibliography is a work in progress to collect and organize studies that use quasi-experimental methods. The literature contained here represents a spectrum of methods of selecting a control group and impact assessment. When available, an abstract of the article is included. RRI expects to continue to update this bibliography.

## 2 Literature

Aleseyed M., Rephann, T., and Isserman A. (1998). The local economic effects of large dam reservoirs: U.S. experience, 1975-95. *Review of Urban and Regional Development Studies* 10(2), 91- 108.

Almus, M. and Czarnitzki, D. (2003). The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany. *Journal of Business & Economic Statistics*, 21, issue 2, p. 226-36.

Angrist, J. and Krueger, A. Empirical Strategies in Labor Economics, "Chapter 23 in O. Ashenfelter and D. Card, eds., *The Handbook of Labor Economics*, Volume III, North Holland, 1999.

Athey, S., and Imbens, G. W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, 74 (2), 431 - 497.

Artz, G., P. Orazem, and D. Otto (2007). Measuring the Impact of Meat Packing and Processing Facilities in Nonmetropolitan Counties: A Difference-in-Differences Approach. *American Journal of Agricultural Economics*, 89 (3), 557-570

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1), 249-275

Bhutta, N. (2009). Regression discontinuity estimates of the effects of the GSE act of 1992. Board of Governors of The Federal Reserve System Finance and Economics Discussion Series Working Paper No 2009-03.

Blundell, R. and Costa Dias, M. (2000). Evaluation Methods for Non- Experimental Data. *Fiscal Studies*, 21(4), 427-468.

Bohm, P. and Lind, H. 1993. Policy evaluation quality: a quasi-experimental study of regional employment subsidies in Sweden. *Regional Science and Urban Economics*. 23 (1) pp. 51-65

- Broder, J., Taylor, T., and McNamara, K. 1992. Quasi-experimental designs for measuring impacts of developmental highways in rural areas. *Southern Journal of Agricultural Economics* 24:199-207
- Card, D. and Krueger, A. B. (1994). A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4).772-793
- Caliendo, M. and Kopeinig, S. (2008), Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22: 31-72
- Cook, T. D., Shadish, W. R., and Wong, V. C. (2008) Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons. *Journal of Policy Analysis and Management*, 27(4), 724-750
- Dehejia, R. (2005). Practical Propensity Score Matching: A reply to Smith and Todd. *Journal of Econometrics*, 125 (1-2), 355-364
- Dehejia, R. H. and Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84(1), 151-161.
- DeVuyst, C. S., Leistriz, F. L., Schepp, A. L.(2003) Socioeconomic Impacts of Agricultural and Non-Agricultural Development in Rural North Dakota. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting Montreal, Canada, July 27-30.
- Galiani, S., Gertler, P., and Schargrotsky, E. (2005). Water for Life: The Impact of the Privatization of Water Services on Child Mortality *Journal of Political Economy*, 113(1), pages 83-120
- Glasmeier, A., and Farrigan, T. (2007). The Economic Impacts of the Prison Development Boom on Persistently Poor Rural Places. *International Regional Science Review*, 30 (3), 274-299.
- Goldstein H. A., Renault C. S. (2004). Contributions of Universities to Regional Economic Development: A Quasi-Experimental Approach. *Regional Studies*, 38(7), 733-746.
- Greenberg M.; Isserman A.; Krueckeberg D.; Lowrie K.; Mayer H.; Simon D.; Sorenson D. (1998). Socioeconomic Impacts of US Nuclear Weapons Facilities - A local-scale Analysis of Savannah River, 1950-1993, *Applied Geography*, 18(2), 101-116.

- Greenstone, M. and Gayer T. (2009). Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management*, 57: 21-44
- Hall, B. and Maffioli, A. 2008. Evaluating the impact of technology development funds in emerging economies: evidence from Latin America. *The European Journal of Development Research*, Taylor and Francis Journals, vol. 20(2), pages 172-198.
- Ham, J. C. and LaLonde, R. J. (2005). Special Issue on Experimental and Nonexperimental Evaluation of Economic Policy and Models. *Journal of Econometrics*, 125(1-2), 1-13.
- Ham, J. C., Li, X., and Reagan, P. B.(2006). Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men. IPER Working Papers No 05.13, Institute of Economic Policy Research, University of Southern California
- Heckman, J., Ichimura H., and Todd P. E. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, 64, 605-654
- Heinrich, C., Maffioli, A., and Vazquez, G. (2010). A primer for applying propensity-score matching. Inter-American Development Bank.
- Hotz, J., Imbens, G., and Mortimer, J. (2004). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125(1-2), 241-270
- Isserman A. and Rephann, T. J. (1995). The Economic Effects of the Appalachian Regional Commission: An Empirical Assessment of 26 Years of Regional Development Planning. *Journal of the American Planning Association*, 61(3), 345-364.
- Isserman A. and Merrifield J. (1987). Quasi-Experimental Control Group Methods for Regional Analysis: An Application to an Energy Boomtown and Growth Pole Theory. *Economic Geography*, 63( 1), 3-19
- Isserman, A., and Merrifield, J. (1982). The use of control groups in evaluating regional economic policy. *Regional Science and Urban Economics* 12(1), 45-58.
- Isserman, A., and Beaumont P. (1989). Quasi-experimental control group methods for the evaluation of regional economic development policy. *Socio-Economic Planning Sciences* 23(1), 39-53.

Jalan, J. and Ravallion, M. (2003). Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching. *Journal of Business & Economic Statistics*, American Statistical Association, vol. 21(1), pages 19-30.

Lechner, M. (1999) The Effects of Enterprise-Related Training in East Germany on Individual Employment and Earnings," *Annales d'Economie et de Statistique*, ENSAE, issue 55-56, pages 97-128.

Lechner, M. (2002) Program Heterogeneity And Propensity Score Matching: An Application To The Evaluation Of Active Labor Market Policies. *The Review of Economics and Statistics*, MIT Press, vol. 84(2), pages 205-220.

Lee David S. and Thomas Lemieux (2009). Regression Discontinuity Designs in Economics. NBER Working Paper No. 14723

Meyer, D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics*, 13(2), 151-161.

Moser, P. (2005). How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs." *American Economic Review*, 95(4): 1214-1236.

Ona L. Y., Hudoyo A., and Freshwater D. (2007). Economic Impact of Hospital Closure on Rural Communities in Three Southern States A Quasi Experimental Approach. *Journal of Regional Analysis and Policy*, 37(2): 155-164

Persson, T., Tabellini, G. and Trebbi, F. (2003), Electoral Rules and Corruption. *Journal of the European Economic Association*, 1: 958-989.

Reed W.R. and Rogers, C. L. (2003). A Study of Quasi-experimental Control Group Methods for Estimating Policy Impacts. *Regional Science and Urban Economics*, 33(1).

Rephann, T. J and Isserman A. (1994). New Highways as Economic Development Tools: An Evaluation Using Quasi- Experimental Matching Methods. *Regional Science and Urban Economics*, 24(6), 723-751

Rephann T. J, Dalton, M., Stair, A., and Isserman A. (1997). Casino Gambling as an Economic Development Strategy. *Tourism Economics*, 3(2), 161-183

Rogers, C. L, and Tao J.L.(2004). Quasi-experimental Analysis of Targeted Economic Development programs: Lessons from Florida. *Economic Development Quarterly*, 18(3), 269-285

Rogers, C. L. and Marshment, R. (2003). Measuring Highway Bypass Impacts on Small Town Business Districts. *RURDS*, 12(3), 250-265

Rosenbaum, P. R. (1989) Optimal Matching for Observational Studies. *Journal of the American Statistical Association*, 84(408), 1024- 1032

Rosenbaum, P. R. and Rubin, D. B. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician* 39 (1), 33-38

Smith, J. A. and Todd, P. E.,(2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125 (1-2), 305-353

Trujillo, A. J., Portillo, J. A., and Vernon, J. A. (2005) The Impact of Subsidized Health Insurance for the Poor: Evaluating the Colombian Experience Using Propensity Score Matching. *International Journal of Health Care Finance and Economics* September 2005, Volume 5, Issue 3, pp 211-239

### 3 Alphabetical Listing of Sources and Abstracts

Alessey M., Rephann T., and Isserman A. (1998). The local economic effects of large dam reservoirs: U.S. experience, 1975-95. *Review of Urban and Regional Development Studies* 10(2), 91- 108.

Dam construction has been an important component of economic development initiatives in the United States. However, few comprehensive ex post empirical studies examine the effects of such projects on local employment and income. This paper employs quasi-experimental control group methods to examine the effects of large dam reservoirs on county income, earnings, population, and employment growth for dams opened in the U.S. during the period 1975-1984. This paper shows that large dam reservoirs have some statistically significant positive effects and tend to stimulate growth. There is considerable variation, however, with dams constructed for flood control purposes and dams located further away from markets and large cities having less of an effect.

Almus, M. and Czarnitzki, D. (2003). The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany. *Journal of Business & Economic Statistics*, 21, issue 2, p. 226-36.

This study analyzes the effects of public R&D policy schemes on the innovation activities of firms in Eastern Germany. The main question in this context is whether public funds stimulate R&D activities or simply crowd out privately financed R&D. Empirically, we investigate the average causal effects of all public R&D schemes in Eastern Germany using a nonparametric matching approach. Compared to the case in which no public financial means are provided, it turns out that firms increase their innovation activities by about four percentage points.

Angrist, J. and Krueger, A. Empirical Strategies in Labor Economics, Chapter 23 in O. Ashenfelter and D. Card, eds., *The Handbook of Labor Economics*, Volume III, North Holland, 1999.

This chapter provides an overview of the methodological and practical issues that arise when estimating causal relationships that are of interest to labor economists. The subject matter includes identification, data collection, and measurement problems. Four identification strategies are discussed, and five empirical examples – the effects of schooling, unions, immigration, military service, and class size – illustrate the methodological points. In discussing each example, we adopt an experimentalist perspective that emphasizes the distinction between variables that have causal effects, control variables, and outcome variables. The chapter also discusses secondary datasets, primary data collection strategies, and administrative data. The section on measurement issues focuses on recent empirical examples, presents a summary of empirical findings on the reliability of key labor market data, and briefly reviews the role of survey sampling weights and the allocation of missing values in empirical research.

Athey, S., and Imbens, G., W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, 74 (2), 431 - 497.

This paper develops a generalization of the widely used difference-in-differences method for evaluating the effects of policy changes. We propose a model that allows the control and treatment groups to have different average benefits from the treatment. The assumptions of the proposed model are invariant to the scaling of the outcome. We provide conditions under which the model is nonparametrically identified and propose an estimator that can be applied using either repeated cross section or panel data. Our approach provides an estimate of the entire counterfactual distribution of outcomes that would have been experienced by the treatment group in the absence of the treatment and likewise for the untreated group in the presence of the treatment. Thus, it enables the evaluation of policy interventions according to criteria such as a mean-



variance trade-off. We also propose methods for inference, showing that our estimator for the average treatment effect is root-N consistent and asymptotically normal. We consider extensions to allow for covariates, discrete dependent variables, and multiple groups and time periods.

Artz, G., P. Orazem, and D. Otto (2007). Measuring the Impact of Meat Packing and Processing Facilities in Nonmetropolitan Counties: A Difference-in-Differences Approach. *American Journal of Agricultural Economics*, 89 (3), 557-570

Considerable controversy exists regarding the costs and benefits of growth in the meat packing and processing industry for rural counties. This study investigates the effects of this industry on social and economic outcomes in nonmetropolitan counties of 23 Midwestern and Southern states from 1990 to 2000. Results suggest that as the meat packing industry's share of a county's total employment and wage bill rises, total employment growth increases. However, employment growth in other sectors slows, as does local wage growth. Industry growth has little impact on local crime rates or on growth of government spending on education, health, or police protection.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1), 249-275

Most papers that employ Differences-in-Differences estimation (DD) use many years of data and focus on serially correlated outcomes but ignore that the resulting standard errors are inconsistent. To illustrate the severity of this issue, we randomly generate placebo laws in state-level data on female wages from the Current Population Survey. For each law, we use OLS to compute the DD estimate of its "effect" as well as the standard error of this estimate. These conventional DD standard errors severely understate the standard deviation of the estimators: we find an "effect" significant at the 5 percent level for up to 45 percent of the placebo interventions. We use Monte Carlo simulations to investigate how well existing methods help solve this problem. Econometric corrections that place a specific parametric form on the time-series process do not perform well. Bootstrap (taking into account the autocorrelation of the data) works well when the number of states is large enough. Two corrections based on asymptotic approximation of the variance-covariance matrix work well for moderate numbers of states and one correction that collapses the time series information into a "pre"- and "post"-period and explicitly takes into account the effective sample size works well even for small numbers of states.

Bhutta, N. (2009). Regression discontinuity estimates of the effects of the GSE act

of 1992. Board of Governors of The Federal Reserve System Finance and Economics Discussion Series Working Paper No 2009-03.

In this paper I estimate the effect of the Underserved Areas Goal (UAG) established under the “GSE Act”, a 1992 law mandating that the housing government-sponsored enterprises Fannie Mae and Freddie Mac help promote credit access and homeownership opportunities for low-income households and in low-income and minority neighborhoods. I identify the goal’s impact by taking advantage of a discontinuity in the census tract eligibility rule. Employing local linear and non-parametric regression discontinuity methods, I find that this goal has had a direct effect on GSE purchasing activity of 3-4% and increases overall GSE-eligible originations by 2-3% on average at the cutoff between 1997 and 2002. Changing eligibility status following the release of Census 2000 data provides another source of variation to identify the UAG’s effect in 2005 and 2006, years of sharply increasing goals levels and years which have contributed heavily to current credit losses. I find that while the UAG affected GSE behavior in 2005 and 2006, GSE risk avoidance limited their response. Unlike previous research, I find no evidence that UAG-induced increases in GSE credit supply crowds-out FHA and subprime lending.

Blundell, R. and Costa Dias, M. (2000). Evaluation Methods for Non- Experimental Data. *Fiscal Studies*, 21(4), 427-468.

This paper presents a review of non-experimental methods for the evaluation of social programs. We consider matching and selection methods and analyses each for cross-section, repeated cross section and longitudinal data. The methods are assessed drawing on evidence from labor market programs in the UK and in the US.

Bohm, P. and Lind, H.. (1993). Policy evaluation quality: a quasi-experimental study of regional employment subsidies in Sweden. *Regional Science and Urban Economics*. 23 (1) pp. 51-65

Policy evaluation methods differ as to the nature of their assumptions about the hypothetical policy-off case. We argue that policy evaluation quality can be raised by using a quasiexperimental method when feasible. In addition, the quality of evaluating policy changes can be raised by designing and publicizing the evaluation method, specified in detail, before any data on target variables are available. These policy evaluation properties characterize the study presented here. It concerns a significant reduction

of payroll taxes in a depressed region of Sweden in order to bolster employment. The evaluation produced clear evidence of a complete lack of employment effects.

Broder, J., Taylor, T., and McNamara, K. (1992). Quasi-experimental designs for measuring impacts of developmental highways in rural areas. *Southern Journal of Agricultural Economics* 24:199-207

Quasi-experimental techniques were developed to provide decision-making tools for documenting the impacts of developmental highways in rural areas. Regression discontinuity analysis (RDA) with limited observations was used to compare economic changes in highway counties to those in non-adjacent control counties. The RDA models found statistically significant changes in population, per capita income, and taxable sales related to highway development. The study found that some counties benefitted from developmental highways, some were unchanged, while some experienced economic decline. RDA models with adjacent controls had better explanatory powers while those with non-adjacent controls were more sensitive to highway-related changes in economic activity. When significant non-highway activities were present, adjacent control models may have understated highway-related impacts, while non-adjacent control models have overstated these impacts. Arguments for using adjacent and non-adjacent experimental designs are discussed.

Card, D. and Krueger, A. B. (1994). A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4).772-793

On April 1, 1992, New Jersey's minimum wage rose from 4.25 to 5.05 per hour. To evaluate the impact of the law we surveyed 4W fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above \$5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment.

Caliendo, M. and Kopeinig, S. (2008), Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22: 31-72

Propensity score matching (PSM) has become a popular approach to estimate causal treatment effects. It is widely applied when evaluating labour market policies, but empirical examples can be found in very diverse fields of study. Once the researcher has

decided to use PSM, he is confronted with a lot of questions regarding its implementation. To begin with, a first decision has to be made concerning the estimation of the propensity score. Following that one has to decide which matching algorithm to choose and determine the region of common support. Subsequently, the matching quality has to be assessed and treatment effects and their standard errors have to be estimated. Furthermore, questions like “what to do if there is choice-based sampling?” or “when to measure effects?” can be important in empirical studies. Finally, one might also want to test the sensitivity of estimated treatment effects with respect to unobserved heterogeneity or failure of the common support condition. Each implementation step involves a lot of decisions and different approaches can be thought of. The aim of this paper is to discuss these implementation issues and give some guidance to researchers who want to use PSM for evaluation purposes.

Cook Thomas D, Shadish William R. and Wong Vivian C.(2008) Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons. *Journal of Policy Analysis and Management*, 27(4), 724-750

This paper analyzes 12 recent within-study comparisons contrasting causal estimates from a randomized experiment with those from an observational study sharing the same treatment group. The aim is to test whether different causal estimates result when a counterfactual group is formed, either with or without random assignment, and when statistical adjustments for selection are made in the group from which random assignment is absent. We identify three studies comparing experiments and regression-discontinuity (RD) studies. They produce quite comparable causal estimates at points around the RD cutoff. We identify three other studies where the quasi-experiment involves careful intact group matching on the pretest. Despite the logical possibility of hidden bias in this instance, all three cases also reproduce their experimental estimates, especially if the match is geographically local. We then identify two studies where the treatment and nonrandomized comparison groups manifestly differ at pretest but where the selection process into treatment is completely or very plausibly known. Here too, experimental results are recreated. Two of the remaining studies result in correspondent experimental and nonexperimental results under some circumstances but not others, while two others produce different experimental and nonexperimental estimates, though in each case the observational study was poorly designed and analyzed. Such evidence is more promising than what was achieved in past within-study comparisons, most involving job training. Reasons for this difference are discussed.

Dehejia, R. (2005). Practical Propensity Score Matching: A reply to Smith and Todd.

Journal of Econometrics, 125 (1-2), 355-364

This paper discusses propensity score matching in the context of Smith and Todd's (Does matching overcome Lalonde's critique of nonexperimental estimators, *J. Econom.*, in press) reanalysis of Dehejia and Wahba (*J. Am. Statist. Assoc.* 97 (1999) 1053; National Bureau of Economics Research working Paper No. 6829, *Rev. Econom. Statist.*, 2002, forthcoming). Propensity score methods require that a separate propensity score specification be estimated for each treatment group-comparison group combination. Furthermore, a researcher should always examine the sensitivity of the estimated treatment effect to small changes in the propensity score specification; this is a useful diagnostic on the quality of the comparison group. When these are borne in mind, propensity score methods are useful in analyzing all of the subsamples of the NSW data considered in Smith and Todd (Does matching overcome Lalonde's critique of nonexperimental estimators, *Journal of Econometrics*, 125 (1-2), 305-353).

Dehejia, R. H. and Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84(1), 151-161.

This paper considers causal inference and sample selection bias in nonexperimental settings in which (i) few units in the non-experimental comparison group are comparable to the treatment units, and (ii) selecting a subset of comparison units similar to the treatment units is difficult because units must be compared across a high-dimensional set of pretreatment characteristics. We discuss the use of propensity score-matching methods, and implement them using data from the National Supported Work experiment. Following LaLonde (1986), we pair the experimental treated units with non-experimental comparison units from the CPS and PSID, and compare the estimates of the treatment effect obtained using our methods to the benchmark results from the experiment. For both comparison groups, we show that the methods succeed in focusing attention on the small subset of the comparison units comparable to the treated units and, hence, in alleviating the bias due to systematic differences between the treated and comparison units.

DeVuyst, C. S., Leistriz, F. L., Schepp, A. L.; Socioeconomic Impacts of Agricultural and Non-Agricultural Development in Rural North Dakota. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting Montreal, Canada, July 27-30, 2003.

Information from the groups of development communities and control communities is

compared and contrasted to discern similarities and differences in the effects of the different types of development initiatives and to develop a set of general principles and recommended actions for community leaders to follow when planning for a new employer.

Galiani, S., Gertler, P., and Schargrodsky, E. (2005). Water for Life: The Impact of the Privatization of Water Services on Child Mortality *Journal of Political Economy*, 113(1), pages 83-120

While most countries are committed to increasing access to safe water and thereby reducing child mortality, there is little consensus on how to actually improve water services. One important proposal under discussion is whether to privatize water provision. In the 1990s Argentina embarked on one of the largest privatization campaigns in the world, including the privatization of local water companies covering approximately 30 percent of the country's municipalities. Using the variation in ownership of water provision across time and space generated by the privatization process, we find that child mortality fell 8 percent in the areas that privatized their water services and that the effect was largest (26 percent) in the poorest areas. We check the robustness of these estimates using cause-specific mortality. While privatization is associated with significant reductions in deaths from infectious and parasitic diseases, it is uncorrelated with deaths from causes unrelated to water conditions.

Glasmeier, A., and Farrigan, T. (2007). The Economic Impacts of the Prison Development Boom on Persistently Poor Rural Places. *International Regional Science Review*, 30 (3), 274-299.

Prison construction experienced explosive growth over the 1980s and 1990s. Many poor rural communities invited prisons into their environs, anticipating jobs and economic development. However, with one notable exception, no ex post empirical studies exist of the economic effects of prison construction on rural counties. Following an extensive review of the literature, this research uses a quasi-experimental control group method to examine the effect of state-run prisons constructed in rural counties between 1985 and 1995 on county earnings by employment sector, population, poverty rate, and degree of economic health. Analysis suggests a limited economic effect on rural places in general, but may have a positive impact on poverty rates in persistently poor rural counties, as measured by diminishing transfer payments and increasing state and local government earnings in places with relatively good economic health. However, there is little evidence that prison impacts were significant enough to foster structural economic change.

Goldstein H. A., Renault C. S. (2004). Contributions of universities to regional economic development: A quasi-experimental approach. *Regional Studies*, 38(7),733-746.

Universities potentially contribute to regional economic development in a number of ways: research, creation of human capital through teaching, technology development and transfer, and co-production of a favorable milieu. We find that the research and technology creation functions generate significant knowledge spillovers that result in enhanced regional economic development that otherwise would not occur. Yet, the magnitude of the contribution that universities' research and technology development activities play is small compared with other factors. We use a quasi-experimental approach to explain variation in change in average earnings per job across 312 metropolitan statistical areas in the USA in 1969-86 and 1986-98.

Greenberg M.; Isserman A.; Krueckeberg D.; Lowrie K.; Mayer H.; Simon D.; Sorenson D. (1998). Socioeconomic impacts of US nuclear weapons facilities - A local-scale analysis of Savannah River, 1950-1993, *Applied Geography*, 18(2), 101-116.

During the 1940s and 1950s, massive facilities were built in the United States to design, construct and test nuclear weapons. What has been the impact of these facilities on the employment, income and population of the surrounding areas? Doubt exists about whether the national security mission was good for the regions where the facilities were built. Using four counties adjacent to the 310-square mile Savannah River nuclear weapons site, we apply a method that estimates the impact by comparing the growth of the counties to a set of counties that were similar prior to the construction of the nuclear weapons facilities. This counterfactual method identified large increases in employment, income and population during the 1950s and the mid-1980s in two counties with weapons facilities. But no benefits and perhaps a negative impact appeared for a third county. The DOE and its contractors are dramatically reducing employment and budgets at Savannah River and other weapons sites. Employment at the site has fallen from over 25&unknown;000 in 1992 to less than 17&unknown;000 in 1996. It may drop to 8000 if no new 'missions' are created. Federal policy toward these regions is discussed in light of these intra-regional variations in economic impact.

Greenstone, M., and Gayer, T. (2009). Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management*, 57: 21-44

This paper argues that an increased application of quasi-experimental and experimental



techniques will improve understanding about core environmental economics questions. This argument is supported by a review of the limitations of associational evidence in assessing causal hypotheses. The paper also discusses the benefits of experiments and quasi-experiments, outlines some quasi-experimental methods, and highlights threats to their validity. It then illustrates the quasi-experimental method by assessing the validity of a quasi-experiment that aims to estimate the impact of the Endangered Species Act on property markets in North Carolina. The paper's larger argument is that greater application of experimental and quasi-experimental techniques can identify efficient policies that increase social welfare

Hall, B. and Maffioli, A. 2008. Evaluating the impact of technology development funds in emerging economies: evidence from Latin America. *The European Journal of Development Research*, Taylor and Francis Journals, vol. 20(2), pages 172-198.

Evaluations of government Technology Development Funds (TDF) in Argentina, Brazil, Chile and Panama are surveyed. All the evaluations were done at the recipient (firm) level using data from innovation surveys, industrial surveys, and administrative records of the granting units, together with quasi-experimental econometric techniques to minimise the effects of any selection bias. TDF effectiveness is found to depend on the financing mechanism used, on the presence of non-financial constraints, on firm-university interaction, and on the characteristics of the target beneficiaries. Four levels of potential impact were considered: R&D input additionality, behavioural additionality, increases in innovative output, and improvements in performance. The evidence suggests that TDF do not crowd out private investment and that they positively affect R&D intensity. In addition, participation in TDF induces a more proactive attitude of beneficiary firms towards innovation activities. However, the analysis does not find much statistically significant impact on patents or new product sales and the evidence on firm performance is mixed, with positive results in terms of firm growth, but little corresponding positive impact on measures of firm productivity, possibly because the horizon over which the evaluation was conducted was too short.

Ham, J. C., and LaLonde, R. J.(2005). Special Issue on Experimental and Nonexperimental Evaluation of Economic Policy and Models. *Journal of Econometrics*, 125(1-2),1-13.

In contrast to cross-sectional analyses, quasi-experimental designs attempt to mimic the conditions for true experiments as closely as possible in field settings (Cook and Campbell 1979; Shadish, Cook, and Campbell 2002). Before introducing statistical



controls for possibly intervening factors, researchers attempt to control their influence by strategically manipulating the study population, time period, and sampling so as to either maximize or eliminate variation in rival factors, thus reducing the probability of omitted variable biases. In studying the economic impacts of institutions of higher education, the regional distribution of universities by size and type would represent the treatment variable. A control group might consist of a set of regions lacking significant knowledge-producing organizations or a different set of regions possessing types of knowledge organizations other than research universities. Additional divisions among groups might be defined on the basis of regional size (proxying agglomeration effects) or other relevant regional characteristics. Although control-group cases cannot be assumed, as with true experiments, to be equivalent to the treatment cases in all respects except for the treatment itself, comparing gain scores—differences in the outcome variables over time, that is, between a pretest and posttest—takes the different starting positions into account.

Ham, J. C., Li, X., and Reagan, P. B.(2006). Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men. IEPR Working Papers No 05.13, University of Southern California, Institute of Economic Policy Research (IEPR)

Our analysis of migration differs from previous research in three important aspects. First, we exploit the confidential geocoding in the NLSY79 to obtain a distance-based measure. Second, we let the effect of migration on wage growth differ by schooling level. Third, we use propensity score matching to measure the effect of migration on the wages of those who move. We develop an economic model and use it to (i) assess the appropriateness of matching as an econometric method for studying migration and (ii) choose the conditioning variables used in the matching procedure. Our data set provides a rich array of variables on which to match. We find a significant effect of migration on the wage growth of college graduates of 10 percent, and a marginally significant effect for high school dropouts of -12 percent. If we use either a measure of migration based on moving across county lines or state lines, the significant effects of migration for college graduates and dropouts disappear.

Heckman, J., Ichimura H., and Todd P. E. (1997) Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, 64, 605-654

This paper considers whether it is possible to devise a nonexperimental procedure for

evaluating a prototypical job training programme. Using rich nonexperimental data, we examine the performance of a two-stage evaluation methodology that (a) estimates the probability that a person participates in a programme and (b) uses the estimated probability in extensions of the classical method of matching. We decompose the conventional measure of programme evaluation bias into several components and find that bias due to selection on unobservables commonly called selection bias in econometrics, is empirically less important than other components, although it is still a sizeable fraction of the estimated programme impact. Matching methods applied to comparison groups located in the same labor markets as participants and administered the same questionnaire eliminate much of the bias as conventionally measured, but the remaining bias is a considerable fraction of experimentally-determined programme impact estimates. We test and reject the identifying assumptions that justify the classical method of matching. We present a nonparametric conditional difference-in-differences extension of the method of matching that is consistent with the classical index-sufficient sample selection model and is not rejected by our tests of identifying assumptions. This estimator is effective in eliminating bias, especially when it is due to temporally invariant omitted variables.

Heinrich, C., Maffioli, A., and Vazquez, G. (2010). A primer for applying propensity-score matching. Inter-American Development Bank.

This document provides a guide to the key aspects of implementing Propensity-Score Matching (PSM) methodology. It summarizes the basic conditions under which PSM can be used to estimate the impact of a program and the data required, presenting examples of PSM applications. It explains how the Conditional Independence Assumption, combined with the Overlap Condition, reduces selection bias when participation in a program is determined by observable characteristics. It also describes different matching algorithms and some tests to assess the quality of the matching.

Hotz, J., Imbens, G., and Mortimer, J. (2004). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125(1-2), 241-270

We investigate the problem of predicting the average effect of a new training program using experiences with previous implementations. There are two principal complications in doing so. First, the population in which the new program will be implemented may differ from the population in which the old program was implemented. Second, the two programs may differ in the mix of their components. With sufficient detail

on characteristics of the two populations and sufficient overlap in their distributions, one may be able to adjust for differences due to the first complication. Dealing with the second difficulty requires data on the exact treatments the individuals received. However even in the presence of differences in the mix of components across training programs comparisons of controls in both populations who were excluded from participating in any of the programs should not be affected. To investigate the empirical importance of these issues, we compare four job training programs implemented in the mid-eighties in different parts of the U.S. We find that adjusting for pre-training earnings and individual characteristics removes most of the differences between control units, but that even after such adjustments, post-training earnings for trainees are not comparable. We surmise that differences in treatment components across training programs are the likely cause, and that more details on the specific services provided by these programs are necessary to predict the effect of future programs. We also conclude that effect heterogeneity, it is essential, even in experimental evaluations of training programs record pre-training earnings and individual characteristics in order to render the extrapolation of the results to different locations more credible.

Isserman A., and Rephann, T. J. (1995). The Economic Effects of the Appalachian Regional Commission: An Empirical Assessment of 26 Years of Regional Development Planning. *Journal of the American Planning Association*, 61(3), 345-364.

The Appalachian Regional Commission (ARC) was formed in 1965 to promote the region's economic development and "to develop comprehensive and coordinated plans and establish planning priorities for the region." For more than a quarter century, it has been a unique federal-state-local planning effort. Although it occupies a secure place in American regional planning history, its continued existence has been far less secure; from the beginning, its strategies and priorities were widely criticized on numerous, diverse grounds. Every year from 1981 through 1988, the Reagan administration attempted to eliminate it. With the recent silver anniversary of the ARC came new books and articles that record its history and achievements. Missing during this time, however, has been a careful empirical analysis of the extent to which the ARC has succeeded in stimulating the Appalachian economy. This paper presents such a study. Using new quasi-experimental control group methods developed by the authors, the paper measures the effects of ARC programs on 391 counties within the region. The major finding is that Appalachia grew significantly faster than did its control group in income, earnings, population, and per capita income. This result also holds for General Appalachia, the poorest sub region.

Isserman A. and J. Merrifield (1987) Quasi-Experimental Control Group Methods for

Regional Analysis: An Application to an Energy Boomtown and Growth Pole Theory. *Economic Geography*, 63( 1), 3-19

University of Texas at San Antonio This paper constitutes an argument for the use of quasi-experimental control group methods as a measurement technique to study economic and spatial structural change. The essence of such methods is the careful identification of a control group—a set of places whose economic development enables measurement of what would have happened in the place under study without the phenomenon or policy being studied. The quasi-experimental approach can be used in economic geography for basic research and planning and policy studies, including measuring the effects of highway investment, airline service, plant closings, tourism activities, dam construction, development initiatives, energy booms, and growth poles. Examples of the last two applications are provided here to illustrate the use and potential of the method.

Isserman, A., and J. Merrifield. 1982. The use of control groups in evaluating regional economic policy. *Regional Science and Urban Economics* 12(1), 45-58.

A major difficulty in measuring the impact of regional policy is the identification of what would have happened in the absence of the policy. Regional science may benefit from the use of quasi-experimental approaches. The specific method presented here combines the concept of control groups with the widely used shift-share framework for evaluating regional policy. Control areas are selected on the basis of their similarity to the aided region in the pre-policy period. The resulting method is sensitive to changing cyclical conditions and other exogeneous factors and traces out the time pattern of impacts.

Isserman, A., and Paul Beaumont. 1989. Quasi-experimental control group methods for the evaluation of regional economic development policy. *Socio-Economic Planning Sciences* 23(1), 39-53.

Abstract Quasi-experimental control group methods can become valuable tools for evaluating public policies and programs that have a spatial dimension. Control groups of places can be used to establish a baseline from which the effect of “treatment” can be inferred. This paper describes the basic approach by presenting an empirical example and then reports on an on-going, multi-year research project intended to make these methods readily usable. The research is focusing on methods for computerizing the selection of control groups and for conducting statistical tests of the significance of the inferred treatment effects.

Jalan, J. and Ravallion, M. (2003). Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching. *Journal of Business & Economic Statistics*, American Statistical Association, vol. 21(1), pages 19-30.

We apply recent advances in propensity-score matching (PSM) to the problem of estimating the distribution of net income gains from an Argentinean workfare program. PSM has a number of attractive features in this context, including the need to allow for heterogeneous impacts, while optimally weighting observed characteristics when forming a comparison group. The average direct gain to the participant is found to be about half the gross wage. Over half of the beneficiaries are in the poorest decile nationally, and 80% are in the poorest quintile. Our PSM estimator is reasonably robust to a number of changes in methodology.

Lechner, M. (1999) The Effects of Enterprise-Related Training in East Germany on Individual Employment and Earnings, " *Annales d'Economie et de Statistique*, ENSAE, issue 55-56, pages 97-128.

The paper studies the returns from enterprise-related continuous vocational training on individual earnings, unemployment probabilities, and other labour market indicators in East Germany after unification. It attempts to solve the intrinsic identification problem of such evaluation problems nonparametrically by using restrictions "produced" by unification as well as by using very informative panel data (GSOEP, 1990-1994). The estimation is performed with nonparametric methods taking account of the panel structure. The results suggest that there are no effects with respect to employment and unemployment probabilities, but that there are large and positive earnings effects.

Lechner, M. (2002) Program Heterogeneity And Propensity Score Matching: An Application To The Evaluation Of Active Labor Market Policies. *The Review of Economics and Statistics*, MIT Press, vol. 84(2), pages 205-220.

This paper addresses microeconomic evaluation by matching methods when the programs under consideration are heterogeneous. Assuming that selection into the different subprograms and the potential outcomes are independent given observable characteristics, estimators based on different propensity scores are compared and applied to the analysis of active labor market policies in the Swiss region of Zurich. Furthermore, the issues of heterogeneous effects and aggregation are addressed. The results suggest that an approach that incorporates the possibility of having multiple programs can be an informative tool in applied work.

Lee David S. and Thomas Lemieux (2009). Regression Discontinuity Designs in Economics. NBER Working Paper No. 14723

This paper provides an introduction and “user guide” to Regression Discontinuity (RD) designs for empirical researchers. It presents the basic theory behind the research design, details when RD is likely to be valid or invalid given economic incentives, explains why it is considered a “quasi-experimental” design, and summarizes different ways (with their advantages and disadvantages) of estimating RD designs and the limitations of interpreting these estimates. Concepts are discussed using examples drawn from the growing body of empirical research using RD.

Meyer D. (1995). Natural and Quasi-Experiments in Economics *Journal of Business & Economic Statistics*, 13(2), 151-161.

Using research designs patterned after randomized experiments, many recent economic studies examine outcome measures for treatment groups and comparison groups that are not randomly assigned. By using variation in explanatory variables generated by changes in state laws, government draft mechanisms, or other means, these studies obtain variation that is readily examined and is plausibly exogenous. This article describes the advantages of these studies and suggests how they can be improved. It also provides aids in judging the validity of inferences that they draw. Design complications such as multiple treatment and comparison groups and multiple pre-intervention or post-intervention observations are advocated.

Moser, P. (2005). How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs." *American Economic Review*, 95(4): 1214-1236.

Studies of innovation have focused on the effects of patent laws on the number of innovations, but have ignored effects on the direction of technological change. This paper introduces a new dataset of close to fifteen thousand innovations at the Crystal Palace World's Fair in 1851 and at the Centennial Exhibition in 1876 to examine the effects of patent laws on the direction of innovation. The paper tests the following argument: if innovative activity is motivated by expected profits, and if the effectiveness of patent protection varies across industries, then innovation in countries without patent laws should focus on industries where alternative mechanisms to protect intellectual property are effective. Analyses of exhibition data for 12 countries in 1851 and 10 countries in 1876 indicate that inventors in countries without patent laws focused on a small set of industries where patents were less important, while innovation in countries

with patent laws appears to be much more diversified. These findings suggest that patents help to determine the direction of technical change and that the adoption of patent laws in countries without such laws may alter existing patterns of comparative advantage across countries.

Ona Lucia Y, Hudoyo Agus, and Freshwater David. (2007). Economic Impact of Hospital Closure on Rural Communities in Three Southern States A Quasi Experimental Approach. *Journal of Regional Analysis and Policy*, 37(2): 155-164

Contradicting the main goals of the Hill-Burton program initiated in the 1940s, many hospitals have since closed in rural communities, mainly during the last two decades. This paper analyzes the economic impact of such hospital closures on rural communities in Georgia, Tennessee, and Texas in the period 1998-2000 by using a quasi-experimental control group method. The essence of this method is the careful identification of a control group – a set of places whose economic development enables measurement of what would have happened in the place under study without the phenomenon or policy being studied. The results indicate that the rural counties that suffered hospital closures did not appear to be affected in economic terms relative to those that did not suffer such a closure.

Persson, T., Tabellini, G. and Trebbi, F. (2003), Electoral Rules and Corruption. *Journal of the European Economic Association*, 1: 958-989.

Is corruption systematically related to electoral rules? Recent theoretical work suggests a positive answer. But little is known about the data. We try to address this lacuna by relating corruption to different features of the electoral system in a sample of about eighty democracies in the 1990s. We exploit the cross-country variation in the data, as well as the time variation arising from recent episodes of electoral reform. The evidence is consistent with the theoretical priors. Larger voting districts—and thus lower barriers to entry—are associated with less corruption, whereas larger shares of candidates elected from party lists—and thus less individual accountability—are associated with more corruption. Individual accountability appears to be most strongly tied to personal ballots in plurality-rule elections, even though open party lists also seem to have some effect. Because different aspects roughly offset each other, a switch from strictly proportional to strictly majoritarian elections only has a small negative effect on corruption.

Reed W. R., and Rogers, C. L. (2003). A study of quasi-experimental control group methods for estimating policy impacts. *Regional Science and Urban Economics*, 33(1).



This study examines the efficacy of quasi-experimental control group (QECG) methods for estimating policy impacts. It establishes that QECG estimators can outperform the conventional regression (CR) estimator when policy adoption is endogenous (nonrandom), the relationship between outcomes and policies is nonlinear, and CR equations do not correctly specify the nonlinear form of the relationship. In the case of perfect matching, QECG methods produce unbiased estimates. In the case of imperfect matching, QECG estimators will be biased. To address this and other issues, we develop a more general QECG estimator that (1) allows control places to be matched to more than one treatment place, and (2) weights observations by the “closeness” of the match. Using Monte Carlo analysis we demonstrate that our estimator substantially improves estimates of policy impacts.

Rephann, T. J., and Isserman A. (1994). New Highways as Economic Development Tools: An Evaluation using Quasi- Experimental Matching Methods. *Regional Science and Urban Economics*, 24(6), 723-751

Stimulating economic growth and development in rural and economically lagging regions is the goal of several federal and state highway programs. This paper examines the effectiveness of highway investment as an economic development tool. A quasi-experimental matching method is used to examine the effects of interstate highways on counties which obtained links during the period 1963-75 or are in close proximity to these newly linked counties. The results show that the beneficiaries of the interstate links in terms of economic growth are interstate counties in close proximity to large cities or having some degree of prior urbanization, such as a city with more than 25,000 residents. Rural interstate and off-interstate counties exhibit few positive effects.

Rephann T. J., Dalton, M., Stair, A., and Isserman A. (1997). Casino Gambling as an Economic Development Strategy. *Tourism Economics*, 3(2), 161-183

Casino gaming has experienced dramatic growth in the United States during the past seven years. Because this growth has occurred recently, there have been few systematic studies of its effects. This paper uses quasi-experimental control group methods to study sixty-eight counties where casinos were opened during the period 1989-1993 and three multi-casino counties. Results show that casino gambling is adopted by economically struggling counties and that it can be a successful development strategy. The effects trickle down to other sectors of the economy, including recipients of income maintenance payments. On the downside, local governments and local workers do not appear to appear to reap the lion’s share of benefits, as much of the income gener-



ated by casinos is dissipated through leakages outside the host county. Finally, some casino types and locations are marginally better than others, but these factors are not prominent determinants of casino effects at this time.

Rogers, C. L., and Tao J.L. (2004) Quasi-experimental analysis of targeted economic development programs: Lessons from Florida. *Economic Development Quarterly*, 18(3), 269-285

The authors highlight three sources of endogeneity bias that typically haunt analyses of local government policies, and offer an empirical methodology for estimating program impacts given such concerns. They investigate Florida's experience with implementing two common targeted economic development policies, community redevelopment areas (CRAs) and enterprise zones (EZs). Developing a simple application decision model as a guide, they find significant differences in policy implementation for small cities compared with larger cities. Florida's small-city program implementation offers a unique opportunity to compare areas that received state-level approval for the programs with all areas that qualified for but did not receive designation using a quasi-experimental framework. In so doing, the authors explicitly address the potential for endogeneity bias caused by program rationing on the part of administrators and by nonrandom targeting of distressed areas. Consistent with existing research, they do not find evidence validating the efficacy of targeted development programs for small cities.

Rogers, C. L., and Marshment, R. (2003). Measuring Highway Bypass Impacts on Small Town Business Districts. *RURDS*, 12(3), 250-265

The literature is inconclusive about the economic impact of highway bypass on small town business districts. Locally imposed sales tax data can be used to construct cross-sectional, time series data for towns too small to be analyzed using U.S. Census data. With these data we examine bypass impacts using standard difference -in-difference methods and supplement the analysis with quasi-experimental analysis. The case study investigates the 1993 bypass of Stonewall, Oklahoma, and a town of approximately 530 people. The analysis indicates the bypass highway had no significant effect on the already declining small town business district. The approach demonstrated in this paper can be used to assess the impacts of wide range of policy treatments at the sub-county level.

Rosenbaum, P. R. (1989) Optimal Matching for Observational Studies. *Journal of the American Statistical Association*, 84(408), 1024- 1032

Matching is a common method of adjustment in observational studies. Currently, matched samples are constructed using greedy heuristics (or “stepwise” procedures) that produce, in general, suboptimal matchings. With respect to a particular criterion, a matched sample is suboptimal if it could be improved by changing the controls assigned to specific treated units, that is, if it could be improved with data at hand. Here, optimal matched samples are obtained using network flow theory. In addition to providing optimal matched-pair samples, this approach yields optimal constructions for several statistical matching problems that have not been studied previously, including the construction of matched samples with multiple controls, with a variable number of controls, and the construction of balanced matched samples that combine features of pair matching and frequency matching. Computational efficiency is discussed. Extensive use is made of ideas from two essentially disjoint literatures, namely statistical matching in observational studies and graph algorithms for matching. The article contains brief reviews of both topics.

Rosenbaum, P. R., and Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician* 39 (1), 33-38

Matched sampling is a method for selecting units from a large reservoir of potential controls to produce a control group of modest size that is similar to a treated group with respect to the distribution of observed covariates. We illustrate the use of multivariate matching methods in an observational study of the effects of prenatal exposure to barbiturates on subsequent psychological development. A key idea is the use of the propensity score as a distinct matching variable

Smith, J. A. and Todd, P. E.,(2005). Does matching overcome LaLonde’s critique of nonexperimental estimators? *Journal of Econometrics*, 125 (1-2), 305-353

This paper applies cross-sectional and longitudinal propensity score matching estimators to data from the National Supported Work (NSW) Demonstration that have been previously analyzed by LaLonde (1986) and Dehejia and Wahba (1999, 2002). We find that estimates of the impact of NSW based on propensity score matching are highly sensitive to both the set of variables included in the scores and the particular analysis sample used in the estimation. Among the estimators we study, the difference-in-differences matching estimator performs the best. We attribute its performance to the fact that it eliminates potential sources of temporally invariant bias present in the NSW data, such as geographic mismatch between participants and nonparticipants and the

use of a dependent variable measured in different ways for the two groups. Our analysis demonstrates that while propensity score matching is a potentially useful econometric tool, it does not represent a general solution to the evaluation problem.

Trujillo, A. J., Portillo, J. A., and Vernon, J. A. (2005) The Impact of Subsidized Health Insurance for the Poor: Evaluating the Colombian Experience Using Propensity Score Matching. *International Journal of Health Care Finance and Economics* September 2005, Volume 5, Issue 3, pp 211-239

This paper evaluates the impact of Colombia's subsidized health insurance program (SUBS) on medical care utilization. Colombia's SUBS program is a demand-side subsidy intended for low-income families, where the screening of beneficiaries takes place in decentralized locations across the country. Due to the self-selection problems associated with non-experimental data, we implement Propensity Score Matching (PSM) methods to measure the impact of this subsidy on medical care utilization. By combining unique household survey data with community and regional data, we are able to compute propensity scores in a way that is consistent with both the local government's decision to offer the subsidy, and with the individual's decision to accept the subsidy. Although the application of PSM using these rich datasets helps to achieve a balance between the treatment and control groups along observable dimensions, we also present instrumental variable estimates to control for the potential endogeneity of program participation. Using both methods, we find that Colombia's subsidized insurance program greatly increased medical care utilization among the country's poor and uninsured. This evidence supports the case for other Latin American countries implementing similar subsidy programs for health insurance for the poor.