

# Empirical Research Design for Public Policy School Students: How to Conduct Policy Evaluations with Difference-in-differences Estimation\*

Hiro Ishise  
Shuhei Kitamura  
Masa Kudamatsu  
Tetsuya Matsubayashi  
Takeshi Murooka

OSIPP, Osaka University

February 21, 2019

This document describes how to write a research proposal for evaluating the impact of some policy with difference-in-differences (DID) estimation. The primary targeted audience are graduate students at public policy schools. We start our discussion with what we think of as a bad research proposal. We then explain why DID estimation provides a better (if not the best) and (importantly) feasible research proposal. Section 3 discusses how to find a good research question that fits DID estimation. Section 4 describes what kind of data and policy you should look for. Section 5 briefly mentions several techniques for DID estimation.<sup>1</sup>

---

\*We thank Miki Kohara for helpful comments on an earlier draft.

<sup>1</sup>Below we assume that you have learned how to conduct DID estimation. For a refresher, read Chapter 5 of Angrist and Pischke (2015). You should read this chapter again and again until you can explain what DID is in your own words. (Another good reference is Chapter 7 of Gertler et al. 2016).

# 1 Bad research proposals

We often see students propose a research question in the form of “What determines X?” where X can be health outcomes, school outcomes, economic outcomes, subject well-being, etc. Then the cross-sectional data from a survey in a particular year is used to run an OLS regression of X on a bunch of regressors. The choice of regressors is made by the literature review.

e.g. The research question: What determines the subjective well-being of people over 65 in China? A survey of Chinese people in 2010 is used to run an OLS regression of subjective well-being on a bunch of regressors. These regressors correspond to what the existing literature suggests as the determinants of happiness.

WE STRONGLY DISCOURAGE YOU from writing such a research proposal. Reasons are three-fold:

1. Lack of originality. We already know from the literature that your regressors do affect the outcome.
2. Omitted variable bias. An OLS estimation with cross-sectional data (unless it’s an RCT) *always* suffers from bias in the estimated coefficients caused by individual-level unobservable factors (e.g. personality, family environment, ethical values, etc.)
3. No concrete policy implications. If data analysis suggests that personal income affects happiness, for example, this finding does not inform policy-makers of any concrete action to take, because your analysis does not suggest what type of income-support policies affect happiness.

The third reason is in particular important for students at public policy schools: they should acquire the skills of data analysis to propose effective policies to policy-makers.<sup>2</sup>

# 2 Why DID?

We believe that the aim of public policy schools in the 21st century is to equip students with the skill set of policy evaluation. In our view, DID estimation constitutes the core of this skill set for three reasons.

---

<sup>2</sup>The lack of concrete policy implications is not necessarily a bad thing *for economics research in general*, however. Economists strive to uncover the mechanism of human behavior, based on which public policies can be devised. If you come up with such a research idea, go ahead. This document intends to help those students with no clue about where to start.

1. DID estimation controls for individual-level unobservable factors as long as they do not change over time. Consequently, estimation results are more credible (if not perfect) as causal impacts than OLS with cross-sectional (observational) data.
2. DID estimation provides a straightforward framework to evaluate the impact of new policies. Consequently, policy implications are easy to derive. There are so many policies out there, which means your research proposal is likely to be original.
3. Among various techniques for policy evaluation, randomized control trials (RCT), regression discontinuity design (RDD), and instrumental variables estimation (IV) are beyond the capacity of most students (and even most faculty members including ourselves). On the other hand, DID estimation is relatively more feasible than these methods. Plus, searching for research questions and datasets with the DID framework in mind may help you discover a policy that can be used for RCT or RDD.

Even if you're not interested in the impacts of public policies, the DID framework is still useful because it can be applied to any causal relationship (the impacts of institutions, culture, social norms, people's behavior etc.).

### 3 Find a good research question

For DID estimation, you need BOTH the **outcomes of interest** and **a policy** to evaluate. If you have trouble finding a good research question, we suggest following either of the two approaches described below. Of course, there are other ways to find a good research question. Ask your supervisor for suggestions.

#### **Outcome-driven**



1. Start with an outcome of your interest.  
  
e.g. Poverty, gender equality, student performance at school, people's health, subjective well-being (i.e. happiness), etc.
2. List up all the possible determinants of your outcome, by surveying the literature, by reading newspapers and magazine articles, and with your educated guesses.

e.g. Subjective well-being is known to be correlated with income, health, marriage, having children, and age among others. See, for example, Ferrer-i-Carbonell and Frijters (2004).

3. Among these possible determinants, find one that can be exogenously changed by a policy.

e.g. One possible determinant of school attendance in developing countries is the distance to the nearest school. Then a policy to build more schools will reduce this distance a lot for some children but not so much for others.

4. Find whether such a policy is enacted in the country of your choice, by reading newspapers etc. If you cannot find one, go back to the previous step.

e.g. Indonesia constructed schools on a massive scale from 1973 to 1978 (Duflo 2001).

5. Now you have a research question: does the chosen policy improve the outcome of your interest?

### Policy-driven

1. Start with a policy of your interest.

e.g. Introduction of a new tax, a reduction in unemployment benefits, the expansion of eligible citizens for child allowance, the construction of schools, reforming the college admission system

2. List up what could be affected by the chosen policy, by surveying the literature, by reading newspapers and magazine articles, and with your educated guesses.

e.g. Subsidizing child care may encourage (or discourage, due to congestion at child care centers) women to participate in labor market. It may change the occupational choice of women. It could also increase the working hours of staff at child care centers.

3. Among these possible outcomes, figure out which ones are more important than others in terms of people's welfare. As long as it's relevant for improving our living standards, economists will accept its importance. Another criterion is originality. Choose the outcomes no one else has looked at.

e.g. Among the possible consequences of installing air-conditioners in public primary schools in Japan, increased profits for air-conditioner manufacturers are less important than pupils' health, attendance and test scores during summer months. The impacts on pupils may have already been investigated but perhaps the impacts on teachers' health may not.

Note: You don't have to focus on one single outcome. Several related outcomes are fine. Examples include student performances (attendance, test scores, the first job income after graduation etc.), people's health (height, weight, whether to see a doctor in the past one year, etc.), firm performances (revenues, profits, loans, investment, etc.)

4. Now you have a research question: does the policy of your interest improve the chosen outcomes?

**Once you find a research question** Talk to your supervisor to see if it's both original and interesting/important. Your supervisor may not be very familiar with the related literature for your research proposal. So do the literature review on your own to check the originality of your research. If you find papers asking the same question, carefully check their methodology and see if yours are better (ask your supervisor on this).<sup>3</sup>

## 4 Find datasets

### 4.1 Panel data for outcomes

DID estimation requires panel data for outcomes. The same unit of observations needs to be observed more than once. Panel data allows you to control for time-invariant unobserved factors. This does not necessarily ensure the estimation of causal impacts, but at least better than cross-sectional regressions. This is the major reason for why you should aim for DID estimation to evaluate public policies.

Consequently, once you nail down your research question, the next job is to find a panel data for your outcomes. There are a few types of panel data.

---

<sup>3</sup>There still exist a bunch of policy evaluation research based on cross-sectional regressions like the one in Section 1. Asking the same research questions as such studies, with DID estimation, will indeed be original and a contribution to the literature.

**Panel surveys** The same unit of observations (households, firms, districts, provinces, countries, etc.) is repeatedly observed, annually, monthly, weekly, daily, or even hourly. Examples include:

- Penn World Table (annual per capita GDP across countries since 1950; see Feenstra et al. 2015)
- Regional GDP data by Gennaioli et al. (2014) (1528 subnational regions in 83 countries since 1950)
- Panel Study of Income Dynamics (US household panel survey since 1968)
- World Input-Output Database (56 industries across 43 countries for 2000-2014; see Timmer et al. 2015)
- Statistik Industri (Indonesian annual census of large and medium-sized manufacturing firms; see Amiti and Konings 2007)

Panel surveys are expensive to implement, and therefore difficult to find for your research. A cross-sectional data of individuals, more easily available, can also be used as a panel data in three ways as described below.

**Repeated cross-sectional surveys** Some cross-sectional surveys are repeatedly conducted in the same locations even though each round interviews different people. You can treat the same age cohort (those born in the same year) across different survey rounds as the unit of observations.<sup>4</sup> See Deaton (1997), section 2.7, for more detail. Examples of repeated cross-section surveys include:

- Living Standard Measurement Surveys (LSMS) by the World Bank, household surveys conducted in various developing countries since the 1980s
- Demographic and Health Surveys (DHS) by USAID, surveys of women of childbearing age in various developing countries since the 1980s

---

<sup>4</sup>For this research design to work, the number of observations in each round of the survey must be very large; otherwise there are only a few individuals in each age cohort from each round, whose average outcome is consequently a very noisy measure of this age cohort's true outcome.

**One cross-sectional survey with the age as the time variable** Even a single cross-sectional survey of individuals can be used as a panel data by seeing the age as time-dimension and the location of residence as the unit of observations. In other words, we annually observe a set of human-beings born in a particular place.

e.g. Duflo (2001) uses a cross-sectional survey of Indonesian men in 1995 and treats this survey as panel data: the age is the time dimension and the district of residence as the location. She compares their income earnings between those reaching the age of primary school by the time a school is constructed in the district and those otherwise. This difference is then compared to the one in the district without any school construction.

**Recall data** Finally, a recall data from a single cross-sectional survey can be used as a panel data. An example is the retrospective fertility survey such as Demographic and Health Surveys. In such surveys, women are interviewed about all of their child births in the past. Then we observe each woman as a mother repeatedly observed whenever she gives birth. See Kudamatsu (2012) as an example of this approach.

## 4.2 Date and location of a policy change

You need to learn *when* and *where* the chosen policy was implemented, to create the treatment indicator variable for your DID estimation. Here are some examples:

- In Card and Krueger (1994) (see also Section 5.2 of Angrist and Pischke 2009), the state of New Jersey raised the minimum wage in April, 1992, while the state of Pennsylvania did not.
- In Richardson and Troost (2009) (see also Section 5.1 of Angrist and Pischke 2015), the central bank provided credits to troubled banks in the southern half of the Mississippi state in 1931, but not in the northern half of the state.

**Date** The policy should have been implemented around the *middle* of the period covered in your panel data. If your data starts after the policy was implemented, you do not observe pre-treatment outcomes. If your data ends before the policy implementation, you do not observe post-treatment outcomes.

e.g. You may be interested in a policy only recently implemented. Obviously the data for post-treatment outcomes are yet to be available. You have to look back at history for the sake of data availability.

**Location** Ideally, a policy should be implemented only in some parts of a country, to conduct difference-in-differences estimation. This way, you can make sure that the treated group and the control group are relatively similar (compared to cross-country comparisons). Countries with decentralized policy-making (United States, China, India, etc.) often see policy changes affecting only parts of the country.

e.g. China is abundant with such sub-national policy changes. Perhaps the most famous is Special Economic Zones to attract foreign direct investment. Wang (2013) evaluates the impact of this policy in the DID estimation.

**Nationwide policy change** On the other hand, countries with centralized policy-making (e.g. Japan) rarely see sub-national policy changes to be exploited for the DID estimation. However, a nationwide policy can still be used for the DID estimation in the following three cases.

1. A nationwide policy change makes every citizen treated, but some citizens were already treated. In this case, the treatment groups are those sub-national areas with less already treated citizens. The rest is the control group. An early example of this approach is Finkelstein (2007).

e.g. Kondo and Shigeoka (2013) look at the introduction of universal health insurance in Japan in 1961. It is a nationwide policy affecting every citizen in Japan. However, its impact should be larger for prefectures with lower health insurance coverage before 1961. Such prefectures form the treatment group while the rest is the control group.

2. A nationwide policy change affects a certain group of citizens only.<sup>5</sup>

e.g. Meyer et al. (1995) examine the impact of increases in benefits for work-related injuries in the U.S. states of Kentucky and Michigan in 1980 and 1982, respectively. While the policy applies to the

---

<sup>5</sup>In this case, be careful to choose the control group. If a policy targets people over 65 years old, for example, every adult under 65 may not be appropriate as a control group because people at their 20s may differ a lot from those over 65. People aged between 60 to 65 may be more comparable to those over 65.



whole of each state, it only affects high-earning individuals (weekly earnings of over \$600). Low-earning individuals are used as a control group in the DID estimation.

3. Theory tells us what types of individuals, firms, or districts are more severely affected by a nationwide policy change. Such types of observations form the treated group while the rest is the control group.

e.g. Baland and Robinson (2008) look at the impact of introducing secret ballots for national elections in Chile in 1958. This is a nationwide policy change. However, their theory predicts that the impact is larger for areas with more landless farmers (because secret ballots prevent landlords to control their votes). So their treatment variable is defined as the period after 1958 interacted with the share of landless farmers over total agricultural workers at the municipality level.

**Your chosen policy may fit RCT or RDD** In the search for policies to be used in DID estimation, you may discover a policy that was randomly assigned by lottery (e.g. corruption audit for Brazilian mayors; see Ferraz and Finan 2008). Or you may discover a policy that can be used for RDD (e.g. unemployment benefit duration changes at a certain age in Germany (Schmieder et al. 2016) and in Austria (Nekoei and Werber 2017)). If this happens to you, congratulations. Start learning estimation techniques for RCT (Duflo et al. 2007; Athey and Imbens 2017) or RDD (Lee and Lemieux 2010; Cattaneo et al. 2018). There's no need to stick to DID estimation.

**Non-policy treatments** The description above can be applied to non-policy treatments. Examples include weather shocks (Dell et al. 2014 for a survey), political events (e.g., Baskaran et al. 2015), and commodity price shocks (e.g., Storeygard 2016).

## 5 DID estimation techniques

Below we just mention several techniques for DID estimation you should be familiar with. These techniques are all standard for any papers published in top journals.

## 5.1 Gather background information on the policy change to check its endogeneity

The policy of your choice may explicitly target certain types of areas (e.g. the areas becoming poorer over time). If so, the parallel trend assumption of DID breaks down. Collect the background information on policy changes in detail. If certain characteristics of areas/citizens affect the probability of treatment by the policy, collect such data to control for (see Sections 5.6 and 5.7).<sup>6</sup>

## 5.2 Equation to estimate

In your paper, you should write down the equation to estimate. For DID estimation, the convention is as follows:

$$y_{it} = \beta T_{it} + \mu_i + \nu_t + \varepsilon_{it} \quad (1)$$

where  $i$  refers to the unit of observations,  $t$  indicates time,  $y_{it}$  is the outcome of interest,  $\mu_i$  the individual fixed effect,  $\nu_t$  the time fixed effect,  $\varepsilon_{it}$  the error term. The treatment indicator,  $T_{it}$ , can be replaced by a continuous variable that measures the intensity of treatment. Add more regressors to equation (1) if needed.

## 5.3 Standard errors estimation

Standard errors should be clustered at the individual level (e.g. households, firms, districts, etc. that are repeatedly observed). This is because the treatment indicator  $T_{it}$  does not vary much within  $i$  (typically 0 up until a particular period and then 1 afterwards) and the outcome is most likely serially correlated, which causes the underestimation of standard errors. See Bertrand et al. (2004) for detail. If there are too few clustering units, clustered standard errors are known to be underestimated (see Angrist and Pischke 2009, section 8.2.3). Use the “wild cluster bootstrap-t” method by Cameron et al. (2008) as a robustness check.

---

<sup>6</sup>A related point is the time lag between the announcement of a new policy and its actual implementation. Policies in Japan, for example, often take quite a long period of time (about five years) from the announcement to the implementation. After the announcement of a new policy, those who will be more affected may start changing their behavior in the anticipation of this policy. Then the parallel trend assumption breaks down. In such a case, using the announcement year as the date of a policy change may be more appropriate.



## 5.4 Event study plot

Estimate

$$y_{it} = \sum_{\tau} \gamma^{\tau} T_i * I_t^{\tau} + \mu_i + \nu_t + \varepsilon_{it} \quad (2)$$

where  $T_i$  is the treatment group indicator,  $I_t^{\tau}$  the indicator of time  $\tau$ . Then plot the estimated  $\gamma^{\tau}$ . These coefficients should be very close to zero up to the beginning of treatment. See Figure 5.2.4 (p. 239) of Angrist and Pischke (2009) for an example.

## 5.5 Concurrent policy changes

A policy change often goes together with other policy changes. The estimated impacts may pick up the effect of such other policies. A couple of ways to deal with this issue:

- If the date and/or the location of policy changes is not exactly the same, control for the indicators of other policy changes. If your results hold, great. If not, frankly admit as a caveat of your research that the estimation results cannot separate the impacts of concurrent policy changes. Honesty is appreciated by other researchers.
- Who benefits the most is likely to be different across policies. By splitting the sample, see what types of households, firms, or districts benefit the most and discuss whether the pattern is consistent with the expected impact of the policy of your interest, not of the others.
- There may be other ways to handle this problem. Learn from papers published in top journals.

## 5.6 Mean comparison

Create a table that compare the means of covariates (time-invariant ones or measured at the beginning of the sample period) between the treatment group and the control group, with t-test results for mean comparison reported.

## 5.7 Interaction of time fixed effects with covariates

As a robustness check, control for the interaction of time fixed effects with covariates (time-invariant ones or measured at the beginning of the sample

period)<sup>7</sup> whose mean values turn out to differ between treatment and control groups (see Section 5.6 above). Such covariates may include region dummies (e.g. Kanto, Kansai, Kyushu, etc. for Japan). For example, de Janvry et al. (2015) control for time fixed effects interacted with those covariates correlated with the treatment group indicator (see Table 1 column 6).

## 5.8 Unit-specific time trends

As a robustness check, control for unit-specific time trends. See pp. 196-201 of Angrist and Pischke (2015) and pp. 238-241 of Angrist and Pischke (2009). This analysis requires at least three periods in your data (otherwise it's collinear with time dummies).

## 5.9 Placebo tests

As a robustness check, replace the dependent variable with an outcome that should not be affected by the policy. If the treatment indicator coefficient is zero for such an outcome, the reader is more convinced of the parallel trend assumption.

## 5.10 Advanced DID estimation

**Triple differences estimation** If treatment is defined not only by area and time but also by certain groups within each area, you can conduct triple differences estimation. With this estimation, it is usually very difficult to come up with omitted variable bias stories. Thus, your estimation results are very convincing. See Yelowitz (1995), which is discussed in Angrist and Pischke (2009), pp. 242-243. Other applications from top 5 journals include Aghion et al. (2008), Bandiera et al. (2017), Cornelissen et al. (2017), Fisman et al. (2017), Muehlenbachs et al. (2015).

**Synthetic Control Methods** Abadie and Gardeazabel (2003) propose an extension of DID estimation for the case where there is only one treated observation. A control observation will be synthesized by taking the weighted average of non-treated observations where weights are obtained in a data-driven way. They apply this method to see the impact of terrorism in Basque on its economic growth. See also Abadie et al. (2010).

---

<sup>7</sup>The time-variant covariates to be interacted should be fixed at the beginning of the sample period, because they may change in response to the policy change. So it's a bad control (pp. 214-217 of Angrist and Pischke 2015).

**Change-in-changes estimation** Athey and Imbens (2006) propose an extension of DID estimation that relaxes the additive assumption on the error term (individual fixed effect plus time fixed effect plus stochastic term).

## 6 Final words

DID estimation is certainly not a panacea. If its common trend assumption breaks down, estimation results are biased. However, compared to the bad research proposals discussed in Section 1, DID gets you inches closer to the unbiased estimate of policy impacts. We believe that public policy school students should at least aim for such policy analysis.

## References

- [1] Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association*, 105(490): 493–505.
- [2] Abadie, Alberto, and Javier Gardeazabal. 2003. “The Economic Costs of Conflict: A Case Study of the Basque Country.” *American Economic Review*, 93(1): 113–132.
- [3] Aghion, Philippe, Robin Burgess, STEPHEN J. REDDING, and Fabrizio Zilibotti. 2008. “The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India.” *American Economic Review*, 98(4): 1397–1412.
- [4] Amiti, Mary, and Jozef Konings. 2007. “Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia.” *American Economic Review*, 97(5): 1611–1638.
- [5] Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- [6] Angrist, Joshua D., and Jörn-Steffen Pischke. 2015. *Mastering ’Metrics: The Path from Cause to Effect*. Princeton University Press.
- [7] Athey, B.Y. Susan, and Guido W. Imbens. 2006. “Identification and Inference in Nonlinear Difference-in-Differences Models.” *Econometrica*, 74(2): 431–497.

- [8] Athey, S., and G.W. Imbens. 2017. “The Econometrics of Randomized Experiments.” *Handbook of Economic Field Experiments*, 1 73–140.
- [9] Baland, Jean-Marie, and James A. Robinson. 2008. “Land and Power: Theory and Evidence from Chile.” *American Economic Review*, 98(5): 1737–1365.
- [10] Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. “Labor Markets and Poverty in Village Economies.” *Quarterly Journal of Economics*, 132(2): 811–870.
- [11] Baskaran, Thushyanthan, Brian Min, and Yogesh Uppal. 2015. “Election Cycles and Electricity Provision: Evidence from a Quasi-Experiment with Indian Special Elections.” *Journal of Public Economics*, 126 64–73.
- [12] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *Quarterly Journal of Economics*, 119 249–275.
- [13] Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *Review of Economics & Statistics*, 90(3): 414–427.
- [14] Card, David, and Alan B. Krueger. 1994. “Minimum Wages and Employment : A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania.” *American Economic Review*, 84(4): 772–793.
- [15] Carpenter, Christopher, and Carlos Dobkin. 2011. “The Minimum Legal Drinking Age and Public Health.” *Journal of Economic Perspectives*, 25(2): 133–156.
- [16] Cattaneo, Matias D., Nicolas Idrobo, and Rocío Titiunik. 2018. “A Practical Introduction to Regression Discontinuity Designs: Volume I.” [sites.google.com/site/rociotitiunik/publications](https://sites.google.com/site/rociotitiunik/publications)
- [17] Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg. 2017. “Peer Effects in the Workplace.” *American Economic Review*, 107(2): 425–456.
- [18] Deaton, Angus. 1997. *The Analysis of Household Surveys*. World Bank Publications.

- [19] de Janvry, Alain, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. 2015. "Delinking Land Rights from Land Use: Certification and Migration in Mexico." *American Economic Review*, 105(10): 3125–3149.
- [20] Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2014. "What Do We Learn from the Weather? The New Climate-Economy Literature." *Journal of Economic Literature*, 52(3): 740–798.
- [21] Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91(4): 795–813.
- [22] Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." in Schultz, T. Paul, and John A. Strauss (Eds.), *Handbook of Development Economics*, vol. 4. Elsevier, pp. 3895–3962.
- [23] Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. "How Important Is Methodology for the Estimates of the Determinants of Happiness?\*" *The Economic Journal*, 114(497): 641–659.
- [24] Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review*, 105(10): 3150–3182.
- [25] Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics*, 123(2): 703–745.
- [26] Finkelstein, A. 2007. "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." *The Quarterly Journal of Economics*, 122(1): 1–37.
- [27] Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. 2017. "Cultural Proximity and Loan Outcomes." *American Economic Review*, 107(2): 457–492.
- [28] Gennaioli, Nicola, Rafael LaPorta, Florencio Lopez-de-Silanes, and Andrei Shleifer. n.d. "Growth in Regions." *Journal of Economic Growth*, 19(3): 259–309.
- [29] Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M.J. Vermeersch. 2016. *Impact Evaluation in Practice, Second Edition*. The World Bank.

- [30] Kondo, Ayako, and Hitoshi Shigeoka. 2013. “Effects of Universal Health Insurance on Health Care Utilization, and Supply-Side Responses: Evidence from Japan.” *Journal of Public Economics*, 99 1–23.
- [31] Kudamatsu, Masayuki. 2012. “Has Democratization Reduced Infant Mortality in Sub-Saharan Africa? Evidence from Micro Data.” *Journal of the European Economic Association*, 10(6): 1294–1317.
- [32] Lee, David S., and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature*, 48(2): 281–355.
- [33] Meyer, Bruce D., W. Kip Viscusi, and David L. Durbin. 1995. “Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment.” *American Economic Review*, 85(3): 322–340.
- [34] Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins. 2015. “The Housing Market Impacts of Shale Gas Development.” *American Economic Review*, 105(12): 3633–3659.
- [35] Nekoei, Arash, and Andrea Weber. 2017. “Does Extending Unemployment Benefits Improve Job Quality?” *American Economic Review*, 107(2): 527–561.
- [36] Richardson, Gary, and William Troost. 2009. “Monetary Intervention Mitigated Banking Panics during the Great Depression: Quasi-Experimental Evidence from a Federal Reserve District Border, 1929–1933.” *Journal of Political Economy*, 117(6): 1031–1073.
- [37] Schmieder, Johannes F., Till von Wachter, and Stefan Bender. 2016. “The Effect of Unemployment Benefits and Nonemployment Durations on Wages.” *American Economic Review*, 106(3): 739–777.
- [38] Storeygard, Adam. 2016. “Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa.” *The Review of Economic Studies*, 83(3): 1263–1295.
- [39] Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries. 2015. “An Illustrated User Guide to the World Input-Output Database: The Case of Global Automotive Production.” *Review of International Economics*, 23(3): 575–605.
- [40] Wang, Jin. 2013. “The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities.” *Journal of Development Economics*, 101 133–147.



- [41] Yelowitz, A. S. 1995. “The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions.” *Quarterly Journal of Economics*, 110(4): 909-39.