Final Project

The Effects of Free Health Insurance in Peru

Professor Erkmen Aslim

SS154 Econometrics

13 December 2018

Abstract

This paper examines the effect of access to health insurance on the probability to receive medical treatment as well as health expenditures in Peru. The institutional setup of the Peruvian health insurance gives a rare opportunity to study the effects of health insurance coverage with the help of a sharp Regression Discontinuity design (RDD.)

I identified statistically significant effects of access to insurance on curative treatment and health expenditures. However, the findings were not robust to the addition of controls and smaller in magnitude than stated in the original paper (Bernal et al., 2017.)[1] I also identified a possible violation of the continuity assumption that undermines the validity of the RDD.

[1]**#studyreplication:** The purpose of this paper is to replicate the findings of Bernal et al., 2017. To do it, I carefully reviewed the study from multiple angles, found relevant literature to gain contextual insights, and used the same econometric methods as well as data. I tried to think critically and not get biased by the results researchers got.

Introduction

There is a proven positive relationship between health insurance coverage and health-related outcomes across a body of studies that use multiple data sources and various econometric approaches from difference-in-differences (Neelsen & O-Donnell, 2017) to instrumental variables (Galárraga et al., 2009). Breen et al. (2001) reported a strong relationship between health insurance and receipt of screening services. A study by Powell-Griner et al. (1999) claims that uninsured people were much less likely to have a regular source of care. Two other papers, Gruber et al. (2014) and Limwattananon (2015) found that the insurance program for poor in Thailand had positive effects on health care utilization and adverse effects on child mortality rates.[2]

Despite these critical findings, many people in developing countries don't have access to health insurance or even basic health care. Peru was one of the states that tried to tackle this problem with the help of its Seguro Integral de Salud (SIS) program that was first introduced in 2001. Thanks to that program, millions of low-income people got access to both preventive and curative services. In 2009, the program was modified so that all individuals who are not formally employed become eligible for free public health insurance given that their welfare index called IFH is below a value of 55. This unique institutional setup gives a rare opportunity to study the effects of health insurance coverage with the help of a sharp regression discontinuity design (RDD).

---

[2] **#sourcequality:** I've conducted a thorough literature review on the effect of health insurance on curative care and health expenditure using only high-quality peer-reviewed primary sources. I also included APA formatted citations where appropriate as well as a list of references.

In this paper, I want to replicate the findings of Bernal et al. (2017) and either confirm or reject their findings that the program has large effects on utilization that are most pronounced for the provision of curative care. They claim that individuals seeing a doctor leads to a higher awareness of health problems and decide to pay out-of-pocket for services that are not covered by insurance.

This issue is of particular relevance because more and more developing countries implement similar policies. However, in many countries like Mexico (Galárraga et al., 2009) and Nicaragua (Thornton et al., 2010), the coverage is not 100% free which results in controversial findings. Since Peruvian insurance is very similar to the Western public health insurance, it provides an excellent opportunity to study whether this insurance type is appropriate for developing countries. The SIS program is also representative of the entire country (45% of the total population in 2011) and can become a role model for countries who plan to introduce health insurance for the poor.

The rest of this paper is organized as follows. The data section describes the dataset I used to build the model as well as summary statistics. The methodology section specifies the model and lists all the underlying assumptions. The results section presents the findings with the help of figures and tables. All the code and the correlation table can be found in the Appendix.

Data

This paper uses cross-sectional data from the Peru National Household Survey, or Encuesta Nacional de Hogares (ENAHO). This survey provides detailed employment and household expenditure data that were gathered to provide information on the living conditions and poverty status of the population. In *Table 1*, you can see the summary statistics of the variables selected for the analysis. They are divided into four categories each of which contains vital information on 1) the eligibility for free insurance, 2) demographic characteristics, 3) health status, and 4) curative care.

For the robustness check, I selected a combination of health-associated factors, such as chronic diseases and accidents, demographics, and access to insurance. I believe that these factors are sufficient to predict the probability of receiving medical treatment. Health variables play an essential role because healthy people don't need to visit doctors. The demographics balance for the intrinsic differences between groups of people (to the left and to the right of the threshold) and also have a significant effect on the number of doctor visits. For example, old people and pregnant females are more likely to need doctors.

For my analysis, I focused on Lima, the capital of Peru. One of the reasons is that I'm using the data from 2011. In this year, the IFH was only applied in this city. Apart from that, Lima is very densely populated. Therefore, there are enough hospitals and doctors so that I can exclude the presence of confounding variables such as a large distance to medical centers or absence of the staff that could affect the demand for health care.

The validity of the RDD relies on those who were just barely treated being the same as those who were just barely not treated. Thus, it makes sense to examine if these groups are similar based on observable variables. *Graphs 1-2* illustrate that the distributions of age and education are similar on both sides of the threshold. The proportion of males to females is also similar on both sides and is around 50%. This covariate balance makes it possible for me to estimate the average treatment effect in the environment in which randomization is unfeasible. Most people, both males and females, have 11 years of education which is equivalent to a high school diploma. The age distribution is slightly skewed to younger people which is representative of developing countries like Peru.

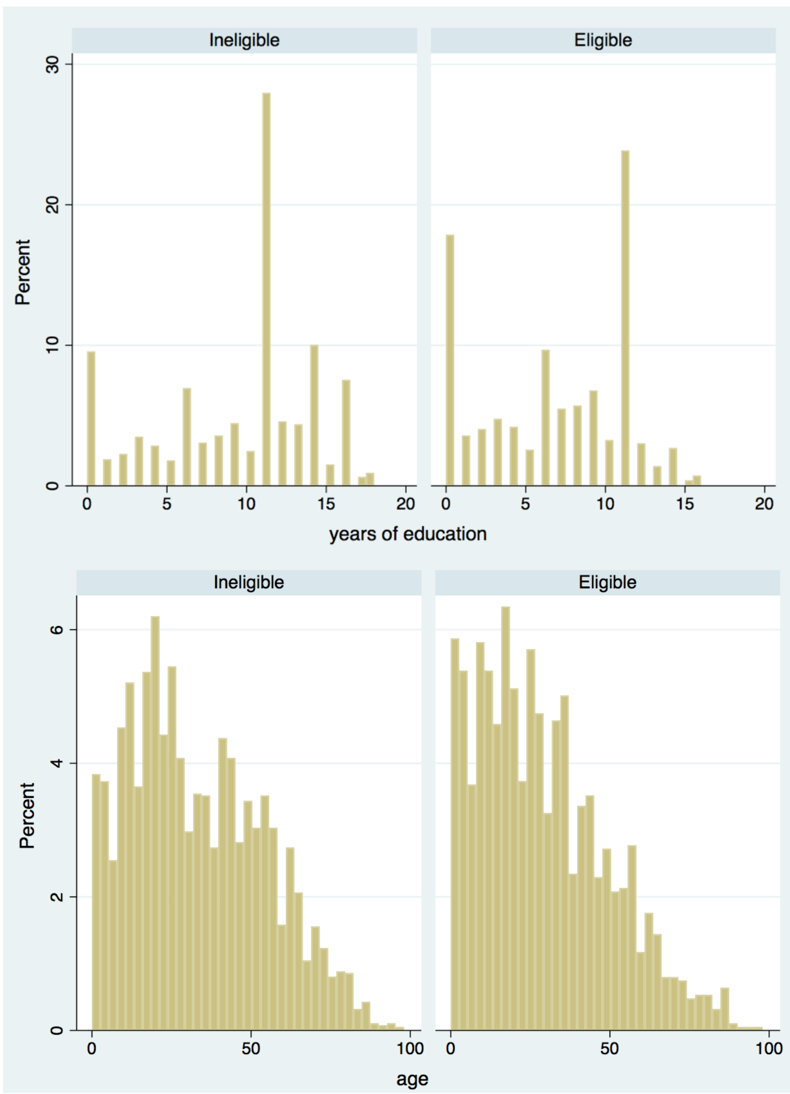

Fig. 1-2. The distribution of education (top) and age (bottom) on different sides of the threshold.

Variable	Mean	SD	Min	Max
IFH minus Threshold	0,61	19,21	-53	45,0
Eligible for IFH	0,43	0,50	0	1
Household Members	4,61	2,10	1	16
Female	0,51	0,50	0	1
Age	33,01	22,25	0	98
Female Head of Household	0,25	0,43	0	1
Education	8,13	4,85	0	18
Chronic Diseases	0,42	0,49	0	1
Symptoms (days)	0,13	1,12	0	30
Illness (days)	0,14	1,10	0	30
Relapse (days)	0,25	2,29	0	30
Accident (days)	0,07	1,12	0	30
Surgery	0.04	0,20	0	30
Curative Care Received	0,25	0,43	0	1
Total Health Expenditures	481	1332	0	22109

Table 1. Summary Statistics. [3]

The first two variables are the running variable and the eligibility criteria. The second set of variables includes demographic factors which serve as controls. The third group is independent

[3]**#dataviz:** I learned how to create tables and graphs that show my findings clearly. Compared to the blind copy-pastes from STATA, this approach gave me a possibility to structure my thoughts and focus on relevant descriptive statistics and estimates. In the RDD outputs, I always included the sample size and standard errors and used asterisks to denote significance level.

predictors of the receipt of curative care and out-of-pocket health expenditures. The last column includes the outcome variables of interest.

Methodology

RD Design

I chose sharp RDD design because it's a simple tool that gives a possibility to estimate causal effects of interventions. By comparing observations that lie closely on either side of the threshold, it is possible to estimate the average treatment effect in environments where randomization is unfeasible.

In this analysis, I used both parametric and non-parametric methods. The `rdrobust` package employs local polynomial and partitioning methods (Calonico et al., 2015.) The major benefit of using non-parametric methods is that they provide estimates based on data closer to the cut-off. This reduces some bias and results in better convergence. However, I will follow the standard approach and do it both ways to show that the estimates are somewhat similar and hence robust.

This paper examines the effect of access to health insurance on the probability to receive medical treatment. The **running variable** is the IFH index that has a sharp eligibility cutoff of 55. In the analysis, I will use *IFH - threshold* to simplify the calculations. The **variable of interest** is *Curative Care* received by individuals. Another dependent variable of interest is *Total Health Expenditure* which will help to check whether access to health insurance increases awareness and the amount of money people spend on their health. I will analyse them independently.

Continuity Assumption

I want to check whether the RDD is applicable in this situation. The key assumption is continuity which means that population average potential outcomes are continuous functions of X at the cutoff. In other words, it means that the expected potential outcomes wouldn't have jumped if there were no treatment. This assumption also implies that there are no unobservable variables that can cause a jump at the cutoff. In our case, the density function would show many households barely qualifying for SIS, to the left of the cutoff, and fewer that do not qualify, to the right of the cutoff if manipulation is present. There is no definitive test for knowing when manipulation occurs, but graphical analysis, such as McCrary test, can help detect when it does.

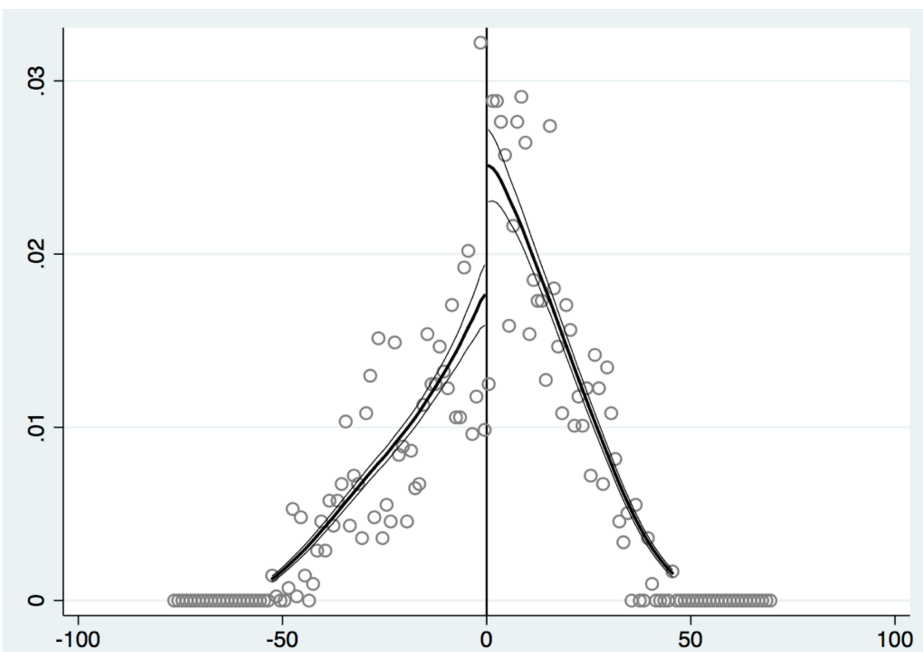

Fig. 3. McCrary test. X axis - IFH, Y axis -density

As indicated above, McCrary test would indicate manipulation in case the density is higher right before the cutoff. This is not the case here. On the contrary, we can see a significant jump in the probability mass right after the cutoff. The continuity assumption is violated here. However, there is no evidence for manipulation from the side of citizens. One possible explanation is that the government wants to save money and makes it extremely hard to satisfy every single eligibility criteria. Alternatively, they want to distribute resources wisely and can only help people who are in extreme need. In any case, only a few people can qualify for free insurance, so that we can see a discontinuity jump at the cutoff point.

Robustness Check

Apart from using parametric and non-parametric approaches, I will also check for robustness by including relevant control variables to see whether the estimates change a lot. Since I'm comparing individuals who are very similar around the cutoff, ideally, the estimates shouldn't be sensitive to the inclusion or exclusion of control variables.

I selected all variables that seemed reasonable factors to seek for curative care. However, the model might have **omitted variables** that could affect the outcome. For example, if people are eligible for insurance but never use it because of their religious views or lack of trust in the health care system. However, this seems to be a rear case and shouldn't influence the outcomes that much.

There is no strong argument in the literature for the **use of quadratic or cubic terms**, so I chose a linear specification. I also checked for **correlation** to avoid including variables that don't add anything to the analysis. Correlation doesn't seem to be a problem here, as illustrated in the Appendix.

Results

First, I will show the relationship between the probability to receive curative care and the IFH index in Fig. 4. Higher index values indicate a higher level of welfare. If the index is below the eligibility threshold, individuals are covered by SIS. The graph illustrates a slight downward jump at zero. The interpretation is that insurance coverage has a positive effect on the probability to receive medical treatment. However, this effect doesn't seem to be significant. I will now use parametric and non-parametric approaches to get numeric estimates.

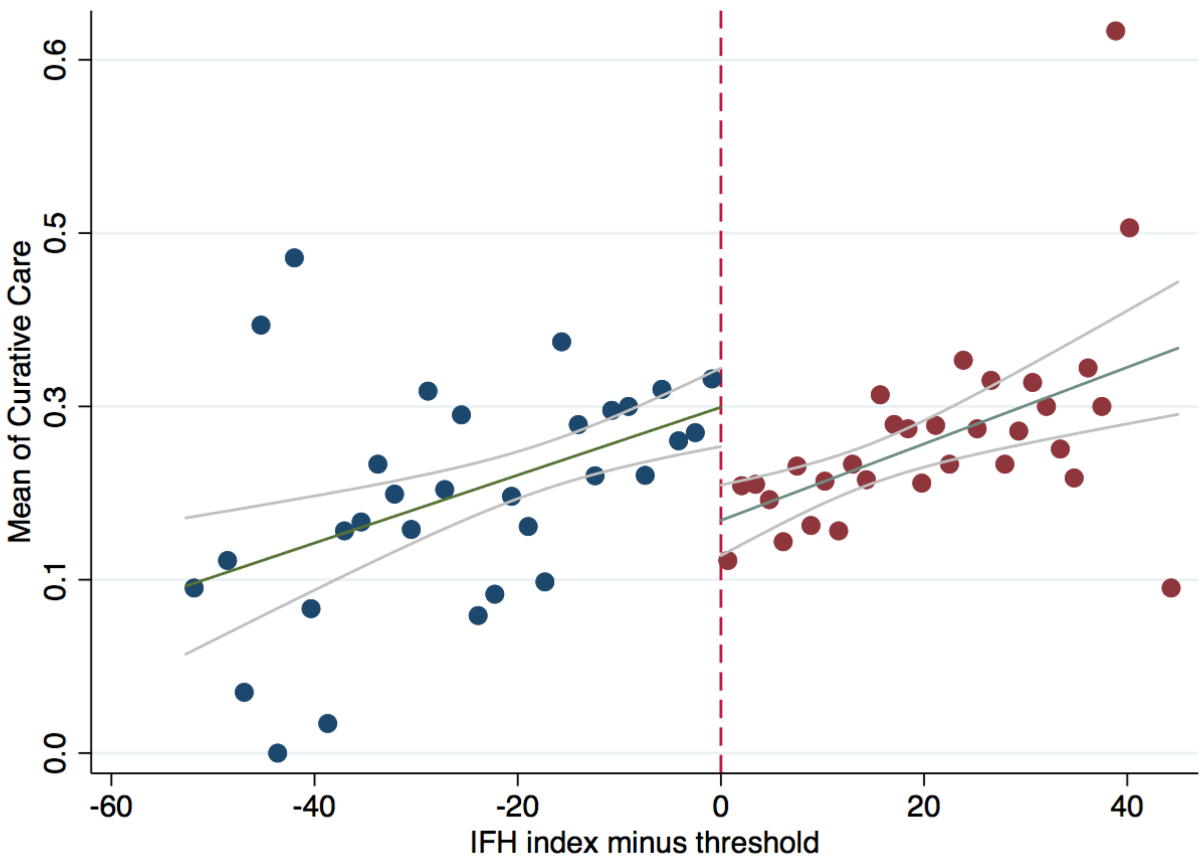

Fig. 4, The RD graph that indicates a discontinuity in the probability of receiving curative care before and after the threshold.

Effects on Curative Care

Before running the regression, I selected the bandwidth using a data-driven approach `rdbwselect` from the `rdrobust` package. The result shows that I should choose individuals with an IFH index that is at most 12 points away from the eligibility threshold. It is also supported by the plot that indicates a lower variance in this range. This estimate is different from the one used in the original paper where they use the bandwidth of 20 from both sides.

After getting the bandwidth, I ran the regression using a non-parametric approach, a parametric approach, and a parametric approach with covariates.

Approach	**Eligibility Estimate**	**Sample Size**	**95% Conf. Interval**	
Non-parametric	0.0991** (0.043)	4161	0.184	0.014
Parametric	0.0978*** (0.023)	4161	0.052	0.143
Parametric with Control Variables	-0.0039 (0.013)	4161	-0.029	0.022

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 2. Regression Discontinuity Estimates for Curative Care

The estimates are robust across different methods and illustrate that being eligible for insurance increases one's chances to receive curative care by 0.09 which is not a lot but statistically significant at 5% confidence level (1% for the parametric approach.) However, the estimate is not robust to the addition of control variables. That means I might have a selection issue around the cutoff given specific covariates. I tried adding each of them separately and saw that the estimate was neither robust nor significant in most of the cases. A potential reason for that is that the effect of being eligible for insurance depends on the health conditions and demographic characteristics (e.g., people with chronic diseases are more likely to use their insurances while healthy people don't go to doctors even if eligible.)

Effects on Health Expenditures

I will now repeat the same analysis to see whether there is an effect of insurance on total health expenditures. This time, the recommended bandwidth is 10 points away from the threshold.

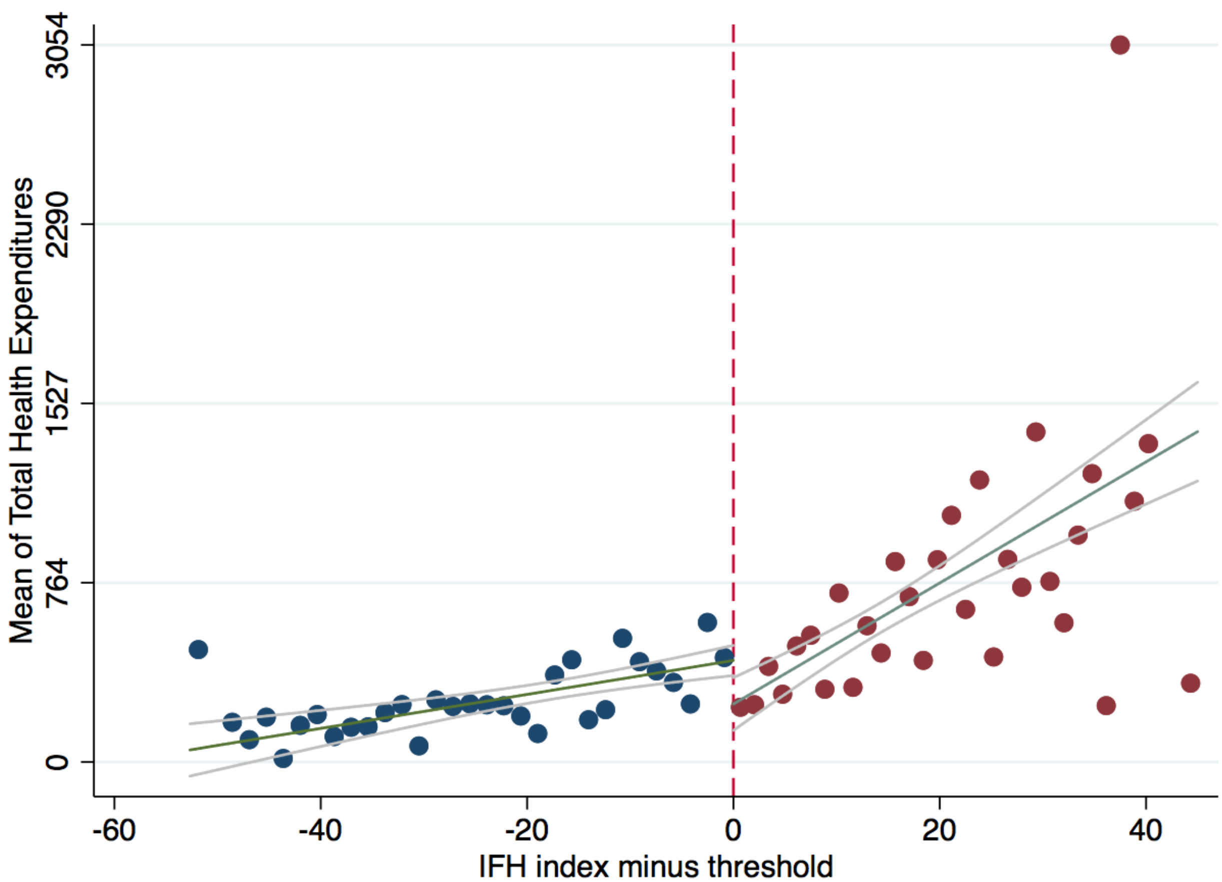

Fig. 5, The RD graph that indicates a discontinuity in the probability of health expenditures before and after the threshold.

Approach	Eligibility Estimate	Sample Size	95% Conf. Interval	
Non-parametric	255.9*** (98.696)	4161	62.45	449.34
Parametric	181.39*** (71.02)	4161	42.15	320.63
Parametric with Control Variables	-209,36*** (40.70)	4161	-289.16	-129.56

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 3. Regression Discontinuity Estimates for Health Expenditures

Compared to the last set of estimates, these estimates are a bit less robust. The parametric and non-parametric estimates have the same sign but a slightly different magnitude. They are statistically significant at 1% level. The inclusion of the covariates has reversed the sign which means that the estimates, similar to the previous case, are not robust to the addition of covariates.

Conclusion

After analyzing the data using parametric and non-parametric Regression Discontinuity design, I can conclude that there is a slight effect of access to insurance on the curative care as well as health expenditures as suggested in the original paper. However, the authors claim that these effects are practically significant and persistent which is rejected by my model. Moreover, the findings are not robust to the addition of control variables which means that insurance might have different effects on different groups of people. I also identified that the continuity assumption might be violated at the cutoff. Overall, I have observed the positive effect of insurance on both outcome variables. However, the scale of this effects as well as their practical significance are questionable and should be studied further. [4]

[4]**#scienceoflearning:** This final project gave me an opportunity to reflect on different aspects of the course and evaluate my progress. I managed to apply various econometric concepts and present my results in a professional manner. This course deepened the introductory knowledge I got in CS112 and equipped me with tools that I will surely use throughout life.

References

Bernal, N., Carpio, M., & Klein, T. (2017). The effects of access to health insurance: Evidence from a regression discontinuity design in Peru. Journal Of Public Economics, 154, 122-136. doi:10.1016/j.jpubeco.2017.08.008

Breen, N., Wagener, D. K., Brown, M. L., Davis, W. W., & Ballard-Barbash, R. (2001). Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. Journal of the National Cancer Institute, 93(22), 1704-1713.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. Journal of the American Statistical Association, 110(512), 1753-1769.

Galárraga, O., Sosa-Rubí, S., Salinas-Rodríguez, A., & Sesma-Vázquez, S. (2009). Health insurance for the poor: impact on catastrophic and out-of-pocket health expenditures in Mexico. The European Journal Of Health Economics, 11(5), 437-447. doi:10.1007/s10198-009-0180-3

Gruber, J., Hendren, N., & Townsend, R. M. (2014). The great equalizer: Health care access and infant mortality in Thailand. American Economic Journal: Applied Economics, 6(1), 91-107.

King, G., Gakidou, E., Imai, K., Lakin, J., Moore, R., & Nall, C. et al. (2009). Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. The Lancet, 373(9673), 1447-1454. doi:10.1016/s0140-6736(09)60239-7

Limwattananon, S., Neelsen, S., O'Donnell, O., Prakongsai, P., Tangcharoensathien, V., Van Doorslaer, E., & Vongmongkol, V. (2015). Universal coverage with supply-side reform:

The impact on medical expenditure risk and utilization in Thailand. Journal of Public Economics, 121, 79-94.

Neelsen, S., & O'donnell, O. (2017). Progressive universalism? The impact of targeted coverage on health care access and expenditures in Peru. Health economics, 26(12), e179-e203.

Powell-Griner, E., Town, M., Nelson, D. E., & Kovar, M. G. (2001, August). National estimates of risks/behaviors: behavioral risk factor surveillance system and the national health interview survey. In Proceedings of the Annual Meeting of the American Statistical Association.

Thornton, R. L., Hatt, L. E., Field, E. M., Islam, M., Solís Diaz, F., & González, M. A. (2010). Social security health insurance for the informal sector in Nicaragua: a randomized evaluation. Health economics, 19(S1), 181-206.

Appendix

Stata Code

```
clear
use "/Users/ritakurban/Downloads/Insurance/data_for_analysis.dta"
keep curative gastosalud2 consulta hospital intervencion hhmujer hospinter consulta_oop formal
consulta_ins consulta_pins medicinas_oop medicinas_ins medicinas_pins Z1 eligibleZ1 telefono
internet cable educ edad mujer lima mieperho sintoma_dias enfermedad_dias recaida_dias
accidente_dias incidente_dias cronica planificacion

* Drop formally employed people since they don't qualify for health insurance
drop if formal==1

* Translate variable names to English
rename Z1 IFH
rename eligibleZ1 eligible
rename hhmujer fem_house
rename edad age
rename mujer female
rename mieperho n_household
rename cronica chronic
rename sintoma_dias symptoms
rename enfermedad_dias illness
rename recaida_dias relapse
rename accidente_dias accident
rename intervencion surgery
rename gastosalud2 expenditures

* Create a Table with Summary Statistics
cd /Users/ritakurban/Downloads
tabstat curative expenditures IFH eligible educ age female n_household chronic symptoms
illness relapse accident surgery fem_house, stat(n mean sd min max) save
return list
matlist r(StatTotal)
matrix results = r(StatTotal)'
putexcel set putexcel2.xlsx, sheet(statistics) modify
putexcel A1 = matrix(results), names nformat(number_d2)

* Histograms to Check Observable Variable Balance
histogram educ, percent by(eligible)
histogram age, percent by(eligible)

* McCrary Test
DCdensity IFH, b(0.5) breakpoint(0) generate(Xj Yj r0 fhat se_fhat)
```

```
* Change Variable Labels for the Cmogram Plots
label var curative "Curative Care"
label var expenditures "Total Health Expenditures"

* Create the Graphs
cmogram curative IFH, cut(0) scatter line(0)  lfitci
cmogram expenditures IFH, cut(0) scatter line(0)  lfitci

* Calculate the Bandwidths
rdbwselect curative IFH
rdbwselect expenditures IFH

* Parametric Approach for Curative
regress curative eligible IFH if IFH<12| IFH>-12
regress curative i.eligible##c.IFH if IFH<12 | IFH>-12
* Include Controls
regress curative eligible female age educ fem_house surgery symptoms illness relapse accident
chronic

* Parametric Approach for Expenditures
regress expenditures eligible IFH if IFH<10| IFH>-10
regress expenditures i.eligible##c.IFH if IFH<10 | IFH>-10
* Include Controls
regress expenditures eligible female age educ fem_house surgery symptoms illness relapse
accident  chronic

* Non-Parametric Approach
rdrobust curative IFH, h(12)
rdrobust expenditures IFH, h(10)
```

Correlation Plot

	visits	paid	IFH	eligible	educ	age	female	n_hous~d	chronic	symptoms	illness	relapse	accident	surgery	fem_ho~e
visits	1.0000														
paid	0.8421	1.0000													
IFH	-0.0139	-0.0534	1.0000												
eligible	0.0360	0.0573	-0.8154	1.0000											
educ	-0.0541	-0.0403	0.3386	-0.2668	1.0000										
age	0.0831	0.0084	0.2159	-0.1840	0.3179	1.0000									
female	0.0656	0.0532	0.0058	-0.0068	-0.0596	0.0424	1.0000								
n_household	-0.0319	0.0071	-0.0740	0.0223	-0.1405	-0.2836	0.0082	1.0000							
chronic	0.1685	0.1020	0.1390	-0.1130	0.0879	0.3923	0.0924	-0.1358	1.0000						
symptoms	0.0814	0.0534	-0.0177	0.0220	-0.0035	0.0508	0.0005	-0.0160	0.0412	1.0000					
illness	0.1529	0.1544	0.0083	0.0059	-0.0055	0.0250	0.0168	-0.0143	0.0340	-0.0035	1.0000				
relapse	0.1308	0.0401	0.0285	-0.0159	-0.0048	0.1119	0.0253	-0.0427	0.1291	0.0197	0.0110	1.0000			
accident	0.0659	0.0621	0.0040	0.0079	0.0176	0.0378	0.0183	-0.0284	0.0400	-0.0044	-0.0080	-0.0067	1.0000		
surgery	0.0409	0.0100	0.0463	-0.0073	0.0567	0.0855	0.0658	-0.0446	0.0724	0.0132	0.0028	0.0696	0.1041	1.0000	
fem_house	0.0047	-0.0079	-0.0064	0.0045	0.0084	0.0249	0.1622	-0.1715	0.0828	0.0113	-0.0118	0.0309	0.0528	0.0063	1.0000