EL COLEGIO DE MÉXICO, A.C.
CENTRO DE ESTUDIOS ECONÓMICOS

"CAUSAL INFERENCE WITH TIME-VARYING CONTINUOUS INTERVENTIONS: EVALUATING THE MEXICAN UNIVERSAL HEALTH INSURANCE PROGRAM SEGURO POPULAR"

TESIS PRESENTADA POR:

CURTIS HUFFMAN ESPINOSA

PARA OPTAR POR EL GRADO DE

DOCTOR EN ECONOMÍA

PROMOCIÓN 2013-2016

DIRECTOR DE TESIS

DR. EDWIN VAN GAMEREN

CIUDAD DE MÉXICO AGOSTO DE 2016

Curtis Huffman Espinosa: *Causal Inference with Time-Varying Continuous Interventions,*
Evaluating the Mexican Universal Health Insurance Program Seguro Popular, © Fall 2016

Committee in charge:

Dr. Edwin van Gameren, Chair
Dr. Raymundo Campos Vázquez
Dr. Laura Juárez González

Location:

Ciudad de México

Time Frame:

August 2016

EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS ECONÓMICOS

CONSTANCIA DE APROBACIÓN

Doctorante: Curtis Huffman Espinosa

Tesis: *Causal Inference with Time-Varying Continuous Interventions: Evaluating the Mexican Universal Health Insurance Program Seguro Popular*

Director de Tesis: Dr. Edwin van Gameren

Aprobada por el Jurado Examinador:

Dr. Edwin van Gameren	Presidente	_____
Dra. Laura Juárez González	Primer Vocal	_____
Dr. Raymundo Campos Vázquez	Vocal Secretario	_____
Dr. Delfino Vargas Chanes	Suplente	_____

Ciudad de México, 10 de agosto de 2016

Causal Inference with Time-Varying Continuous Interventions: Evaluating the Mexican Universal Health Insurance Program Seguro Popular

by

Curtis Huffman Espinosa

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

at the

Centro de Estudios Económicos
of
El Colegio de México, A.C.

Committee in charge:

Dr. Edwin van Gameren, Chair
Dr. Raymundo Campos Vázquez
Dr. Laura Juárez González

Fall 2016

The dissertation of Curtis Huffman Espinosa, titled Causal Inference with Time-Varying Continuous Interventions: Evaluating the Mexican Universal Health Insurance Program Seguro Popular, is approved:

_____

Chair Date

_____

Date

_____

Date

El Colegio de México, A.C.

Causal Inference with Time-Varying Continuous Interventions: Evaluating the Mexican Universal Health Insurance Program Seguro Popular

A Margarita y Dennis, por aclarar el camino.

Abstract

Causal Inference with Time-Varying Continuous Interventions: Evaluating the Mexican Universal Health Insurance Program Seguro Popular

by

Curtis Huffman Espinosa
Doctor of Philosophy in Economics

El Colegio de México, A.C.

Dr. Edwin van Gameren, Chair

In Mexico, over 50 million people without health insurance in 2001, now have a state-protected health coverage via the public insurance system Seguro Popular (SP). Here we assess the impact of SP on the human and material resources needed to meet the new demand for health services. This has required new procedures, which in turn motivated the development of a robust estimator, available to the community as a do-file for the statistical software Stata®. In Chapter 1 we analyze the implications of different causal assumptions in evaluating SP. There we show the need to push forward Propensity Score methods so as to accommodate time-varying continuous treatments. In Chapter 2, we present a new semiparametric procedure that allows researchers to distill causal quantities in these contexts. This is accomplished by bringing together the literature on continuous and dynamic treatments. The proposed procedure allows researchers to estimate Mean and Quantile Dose-Response Functions by applying local regression methods to appropriately weighted samples that control for time-dependent confounding. It is in this Chapter that we estimate and discuss the impact of SP. Given the complexity behind the procedure discussed in Chapter 2, in Chapter 3 we present a robust estimator that facilitates causal analysis in dynamic settings with continuous treatments. All of our estimates sug-

gest that, on the average, SP has effectively had a positive impact on the Mexican Ministry of Health's resources. However, we find compelling quantitative evidence that the program has proven most helpful in less vulnerable territories, leaving behind those in greater need.

Contents

List of Figures **v**

List of Tables **vii**

1 Seguro Popular and the Supply of Health Care Services **1**
1.1 Introduction 2
1.2 Evaluating Seguro Popular 3
1.3 Data 7
1.4 Empirical Strategy 9
1.5 Results 16
1.6 Conclusions and Discussion 33

2 Time-Varying Continuous Interventions **40**
2.1 Introduction 41
2.2 Causal Inference 43
2.3 Estimating the Stabilized Weights 49
2.4 Mean and Quantile Dose-Response Functions for Time-varying Continuous Interventions 53
2.5 Estimating the Effect of Seguro Popular on the Supply of Healthcare Services 54
2.6 Conclusions 82

3 Robust Estimation of Inverse Probability Weights **84**
3.1 Introduction 85
3.2 Marginal Structural Models 87
3.3 Correlation Breaking IPTW 90
3.4 Simulation Studies 94
3.5 Empirical Application 97
3.6 Conclusions 98

Final Comments **103**

Resumen (Español) **104**

Bibliography **121**

A Tables and Derivations **128**
A.1 Tables from Chapter 1 128
A.2 Derivation and Tables from Chapter 3 153

B Programming code (Stata®) **166**

List of Figures

1.1 Seguro Popular Rollout 4
1.2 National Evolution of Key Resources and SP Coverage 8
1.3 Causal Diagram Implicit in Observational Studies with Time-independent Confounding 10
1.4 Distribution of SP Coverage Across Sanitary Jurisdictions in 2009 11
1.5 Causal Diagram Implicit in Median Dichotomized Continuous Treatment 13
1.6 Descriptive Statistics of the Sanitary Jurisdictions 18
1.7 Average Treatment Effects of Dichotomized SP coverage on Sanitary Jurisdictions by Matching Algorithm 23
1.8 Covariate-Balance for SP coverage with Balanced Samples with respect to Dichotomized Treatment on Sanitary Jurisdictions 25
1.9 Pre-Treatment-Covariate-Balance for SP coverage on Sanitary Jurisdictions by Weighting Algorithm 28
1.10 Average Treatment Effects of SP coverage on Sanitary Jurisdictions by Balancing Algorithm 30
1.11 Quantile Treatment Effects of SP coverage on Sanitary Jurisdictions by Balancing Algorithm 32
1.12 Interaction Effects of SP coverage with Pre-Treatment Covariates on Sanitary Jurisdictions by Balancing Algorithm 37
1.13 Causal Diagram Illustrating Time-dependence 39

2.1 Stabilized Weights Over the Years 66
2.2 Increment in the Relative Number of Doctors' Offices 68
2.3 Increment in the Relative Number of Physicians in day-to-day Contact with Patients 69
2.4 Increment in the Relative Number of Nurses in day-to-day Contact with Patients 70
2.5 Relative Number of Human and Material Resources 2001 72

2.6 Expected Increment in the Relative Number of Doctors' Offices due to Full Coverage of SP 73
2.7 Expected Increment in the Relative Number of Physicians in day-to-day Contact with Patients due to Full Coverage of SP 74
2.8 Expected Increment in the Relative Number of Nurses in day-to-day Contact with Patients due to Full Coverage of SP 75
2.9 Increment Incidence Curve for the Relative Number of Doctor's Offices, 2001-2013 78
2.10 Increment Incidence Curve for the Relative Number of Physicians in day-to-day Contact with Patients, 2001-2013 79
2.11 Increment Incidence Curve for the Relative Number of Nurses in day-to-day Contact with Patients, 2001-2013 80
2.12 Inequality in SSA Human and Material Resources, 2001-2013 81

3.1 Data Generating Process in Simulation Study 94
3.2 Covariate-Balance for Time-varying SP coverage on Sanitary Jurisdictions 99
3.3 Average Treatment Effects of SP coverage on Sanitary Jurisdictions . . . 101

E.1 Gasto ejercido por el Seguro Popular 105
E.2 Establecimientos de Salud 106
E.3 Recursos 107
E.4 Egresos Hospitalarios 110
E.5 Densidad de recursos 112
E.6 Impacto del SP por JS 116

List of Tables

1.1 Covariate-Balance for Dichotomous Treatment on Sanitary Jurisdictions . 19

2.1 Descriptive Statistics of the Whole Sample 57
2.2 Descriptive Statistics of the Trimmed Sample 60
2.3 The Change in History-Adjusted Balance between the Weighted and Unweighted Data 64

3.1 Performance of Probability Weights Estimation Methods 97

A.1 Descriptive Statistics of the Sanitary Jurisdictions 129
A.2 Average Treatment Effects for Dichotomized SP coverage on Sanitary Jurisdictions 131
A.3 Covariate-Balance for SP coverage with Balanced Samples with respect to Dichotomized Treatment on Sanitary Jurisdictions 133
A.4 Covariate-Balance for SP coverage on Sanitary Jurisdictions 140
A.5 Average Treatment Effect of SP coverage on Sanitary Jurisdictions 144
A.6 Quantile Treatment Effect of SP coverage on Sanitary Jurisdictions 147
A.7 Interaction Treatment Effect of SP coverage with Pre-treatment Covariates on Sanitary Jurisdictions 150
A.8 Covariate-Balance for Time-varying SP coverage on Sanitary Jurisdictions 161
A.9 Average Treatment Effect of SP coverage on Sanitary Jurisdictions 164

Acronyms

ATE	Average Treatment Effect
CBPS	Covariate Balancing Propensity Score
CONAPO	Consejo Nacional de Población
DGP	Data-Generating Process
DRF	Dose-Response Function
GLM	Generalized Linear Model
GM	Genetic Matching
GMM	Generalized Method of Moments
GPS	Generalized Propensity Score
IMSS	Intituto Mexicano del Seguro Social
IPTW	Inverse Probability of Treatment Weights
ISSSTE	Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado
LGS	Ley General de Salud –General Health Law–
MSM	Marginal Structural Model

OLS Ordinary Least Squares

PEMEX Petróleos Mexicanos

PS Propensity Score

QTE Quantile Treatment Effect

REPSS Regímenes Estatales para la Protección Social en Salud –State Regimes for Social Protection in Health–

RMSE Root Mean Squared Error

SINAIS Sistema Nacional de Información en Salud –National Health Information System–

SJ Sanitary Jurisdiction

SP Seguro Popular

SPSS Sistema de Protección Social en Salud –System of Social Protection in Health–

SSA Secretaría de Salud –Mexican Ministry of Health–

TSCS Time-series Cross-Sectional

UNEMES Unidades de Especialidades Médicas

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Edwin van Gameren for the continuous support of my Ph.D study and related research. Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Raymundo Campos Vázquez and Prof.Carlos Chiapa Labastida, for their insightful comments which incented me to widen my research from various perspectives. This thesis would not have been possible without the financial support from CONACYT and El Colegio de México, A.C.

Chapter 1

Seguro Popular and the Supply of Health Care Services

In this Chapter we assume different degrees of homogeneity in Seguro Popular's intervention and analyze their implications for its evaluation. All of our estimates suggest that, on the average, the Seguro Popular program has effectively had a positive impact on Secretaría de Salud's (the Mexican Ministry of Health) resources. However, quantile and interaction treatment effects also suggest that the program may be leaving behind some of the most vulnerable geographical areas.

1.1 Introduction

By the end of 2012 Mexico reached a truly immense landmark in its journey towards universal health insurance coverage. Some 50 million people uninsured in 2001, basically all informal workers and their families, now have a state-protected health coverage via the public insurance system Seguro Popular (SP). This is without any doubt the most important financial effort to provide health insurance for the uninsured the country has ever made since the creation of its social security system in 1943. However, it would be naive to assume that achieving universal health coverage, as it has been referred to [31], has automatically been translated into the human and material resources needed to meet the new demand for health services. In this paper we examine the degree to which this has happened.

We make two distinct contributions. First, although the goal of the SP is to increase access to health care, surprisingly few papers have examined its impact on the service delivery in general, and on the provision for human and material resources involved in the supply of health care services in particular; that is, on the supply side of the health care services. So far, the evaluations of SP have mainly focused on the effects on out-of-pocket health payments, catastrophic expenditures, health, and labor force participation –a notable exception is a paper by Bosch and Campos [5]. However, since the program's funding goes without strict earmarking, whether these financial resources managed to translate into resources associated with the provision for healthcare services is an important matter that deserves attention.

Secondly, in working out the implications of turning to a number of simplifying statistical and causal assumptions (dichotomizing an otherwise continuous time-varying treatment) in order to apply existing propensity score methods, we show the need for a further extension to the time-varying continuous context. The pursuit of such research agenda will allow researchers to incorporate causal assumptions otherwise very hard to work with, given the current state of knowledge we have.

All of our estimates suggest that, on the average, more human and material resources followed geographical areas with higher degrees of SP coverage. Even more so, analyzing the heterogeneity behind the average treatment effects, quantile and interaction treatment effects indicate that the SP program has had distributional effects on said resources, leaving behind some of the most vulnerable parts of the country. Whether this is an unintended effect of the program or not, its inequality implications need to be addressed openly, for it means that a different kind of public health policy is needed to attend to the most vulnerable part of the population in the country.

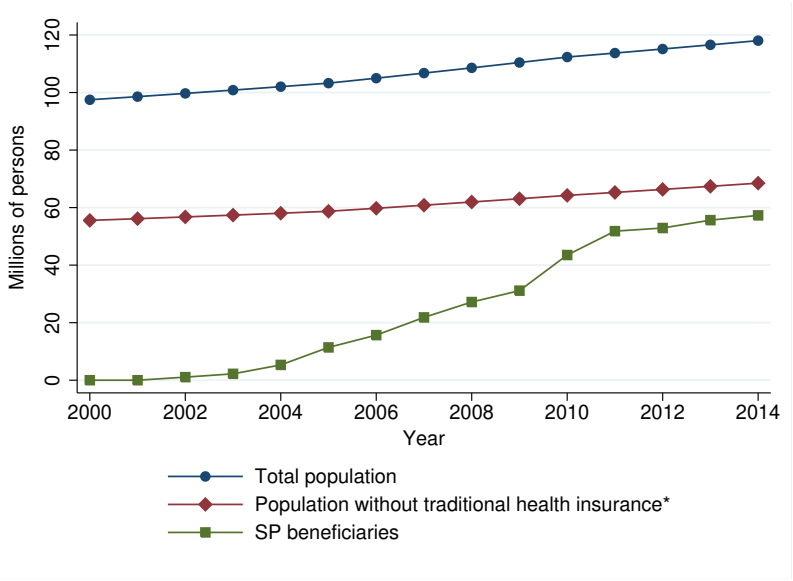
The paper proceeds as follows. Section 1.2 presents a short institutional background, the main features of the program and reviews the evaluation literature related to SP. Section 1.3 describes the data, and Section 1.4 presents the empirical strategies. In Section 1.5 we present the corresponding estimates, and finally in Section 1.6 we conclude with a brief discussion of our results.

1.2 Evaluating Seguro Popular

Seguro Popular (Popular Health Insurance, SP) was launched as a pilot program in 2002 and remained so until the end of 2003. At that time, it covered roughly 2.2 million people previously uninsured in a total population of about 100 million, of which over 50 percent lacked any health insurance coverage, leaving them effectively outside the social health protection system. By that same time, the new Ley General de Salud –General Health Law– (LGS) launched the Sistema de Protección Social en Salud –System of Social Protection in Health– (SPSS) and just over 10 years later the SP covers more than 55 million people.

Enrollment into the System is voluntary, and is granted to all legal residents in Mexican territory who lack health insurance, ascertained with the mere declaration of the applicant. Figure 1.1 shows the evolution of this huge effort to provide health insurance for the uninsured.

Figure 1.1: Seguro Popular Rollout



*Traditional health insurance here refers to the one provided by public institutions (work-related) such as IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy

Source: Personal elaboration based on data from the Census on Population and Housing and administrative records from Seguro Popular

It's noteworthy that the LGS specified that the Federal Government and the states were to share the responsibility, being the later responsible for managing the resources allocated by the Federation for the service delivery in general. The funding of SP has a tripartite structure: a social contribution from the federal government, covered annually by the federal government directly proportional to the number of beneficiaries (the main source of funding); a solidarity contribution, covered both from the federal government and states according to health indicators such as infant and adult mortality; and a family contribution, a fee introduced to replace out-of-pocket payments made at the time of the delivery of services, although families can be classified into a non-contributory regimen based on the household's income. By 2012, the beneficiaries had access to a package of 284 health services and interventions listed in the Catálogo Universal de Servicios Esenciales de Salud (Universal List of Essential Health Services, CAUSES).

Previous Impact Evaluations

Key findings of the many evaluations of SP include that SP affiliates see reduced their out-of-pocket health expenditures [50, 15, 3, 28, 18, 37, 2] and catastrophic health expenses [20]. Increments in the use of services have also been reported [32, 50, 15, 37], as well as positive effects not just on the self-reported health of the beneficiaries [50], but also in some biomarkers [37].

Another strand of the evaluation literature of SP has focused on its impact on the labor market, as the program might be creating perverse incentives to work in the informal sector. However, here the evidence is mixed. Some studies find negligible or no effects [11, 36, 1, 6], while others find that the SP actually has a negative effect on the creation of formal jobs, especially in small and medium sized firms [5].

Surprisingly, practically no attention has been given to the effects that the expansion of this nationwide health insurance program might have on the service delivery in general; that is, on the supply side of the health care services. So far, only [5] have tackled this issue providing evidence that the SP program has had a positive impact on the number of

physicians, nurses and clinics in the Mexican municipalities. These authors followed a difference-in-differences approach conceptualizing the implementation of the SP program as an event that occurs in the municipalities once more than 10 individuals have been affiliated.

Of course, the chosen unit of analysis and the way researchers operationalize the treatment play an important role in the nature of the findings. In this regard, two aspects of Bosch and Campos' [5] strategy are worth mentioning. First, it's not clear that the municipality is the natural unit of analysis regarding human and material health resources. In Mexico, the basic regional administrative unit in charge of the operation of healthcare services and programs is the Sanitary Jurisdiction (SJ). Mexican SJs are, in general, comprised of several municipalities which vary greatly both in area and population.[1] Secondly, more than 10 individuals is hardly a clear sign of SP coverage across such diverse geographic areas as the Mexican municipalities.

In the remainder of the paper we contribute to the evaluation literature of SP going deeper in assessing heterogeneous effects of SP on basic resources for the provision for healthcare services. It is a well known fact that every impact evaluation incorporates in its empirical strategy both statistical and causal assumptions about the data-generating process [38], and that imposing different assumptions leads to different causal models and quantities to be estimated. On our part, we focus on the SJ as unit of analysis, assessing the implications of a set of different statistical and causal assumptions. Whenever possible, we test the compatibility of these assumptions with the data. However, not all assumptions lend themselves to receive empirical scrutiny. Most importantly, regarding our estimates, everything will be predicated on the validity of the no-confounding assumption, also known as selection-on-observables or ignorability. In other words, that we have identified and measured all possible variables whose effect may be confounded with that of the SP coverage.

[1] Details in Section 1.3.

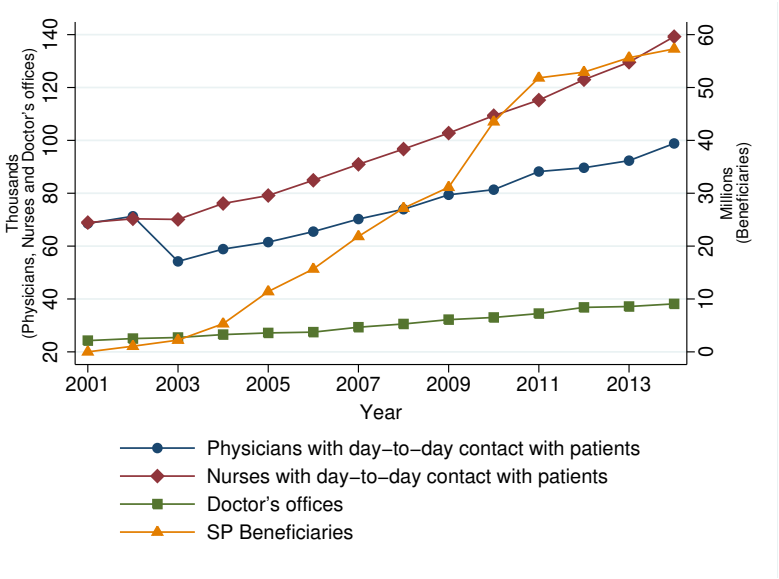
1.3 Data

We have merged the administrative records of SP containing the number of families and individuals affiliated from 2002-2013 with the federal records of infrastructure and human resources employed by the Secretaría de Salud –Mexican Ministry of Health– (SSA). These are the state's primary resources used in providing health care services to the population not insured by any of the traditional public institutions. The SSA data is available at a yearly frequency disaggregated at the health establishment (service outlet) level, and nowadays it's possible to retrieve this data from 2001 to 2014 from the Sistema Nacional de Información en Salud –National Health Information System– (SINAIS) web site: http://www.sinais.salud.gob.mx/.

Figure 1.2 shows the evolution of key resources involved in the supply of health care services along with the coverage of the SP program. There we can see that the increased coverage of the SP program seems to be associated with a growing number of physicians and nurses in day-to-day contact with patients (providing clinical care) as well as with the number of doctors' offices in its national aggregate. In the rest of the paper we shall focus on these three variables since they seem most fundamental in the provision for healthcare services. Unfortunately, the SINAIS database does not go any further back than 2001 and we cannot appreciate properly whether these trends correspond to a change in the historical tendency of the variables. It is, however, from this apparent association that we strive to disentangle a causality relationship.

Only 19% of the municipalities in Mexico register an SSA health establishment, which hardly makes it the best unit of analysis for our purposes. In all our estimates we use the sanitary jurisdiction as unit of analysis and aggregate the data accordingly. There are 242 SJs in Mexico and none of them cross states. However, 9 out 2,457 municipalities cross sanitary jurisdictions. For lack of better data access, in aggregating the data we have treated these municipalities as a sanitary jurisdiction in themselves. Leaving us with a panel of 233 units across 13 years.

Figure 1.2: National Evolution of Key Resources and SP Coverage



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

As noted before, the sanitary jurisdiction is the basic regional administrative unit of the SSA in charge of the operation of healthcare services. Also important to note is that, given the variation in size and population across SJs, we focus on the density of SSA's material and human resources, not in their absolute numbers, in all cases relative to the population that lack labor-related health insurance coverage provided by public institutions such as Intituto Mexicano del Seguro Social (IMSS), Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE), Petróleos Mexicanos (PEMEX), the Ministry of Defense or the Navy. The same reference population is used for the SP coverage, only this time expressed as a proportion.

1.4 Empirical Strategy

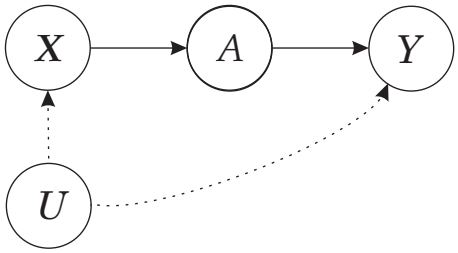
In view of the possibility to draw on the SINAIS database for annual series of SJ's characteristics, ranging from a year prior to the implementation of SP –since the SP program started as a pilot during 2002– to 2014, the combination of difference-in-differences, which has the additional advantage of controlling for individual permanent effects, with Propensity Score (PS) methods seem best suited to assess whether human and material resources followed the new entitlement to meet the new potential demand for services. Under an unconfoundedness assumption, also known as the assumption of selection on observables, these methods allow removing all biases in comparisons in assessing the treatment effect by adjusting for differences in a set of covariates.

We use a set of 7 variables. We focus on the density (relative number) of doctors' offices, staffed and non-staffed hospital beds, physicians and nurses, both with and without day-to-day contact with patients (with and without clinical duties). In particular, at any given point in time, we look for the effects SP has had on the change (increment) in the density of these human and material resources, conditioning on their levels prior to the exposure to SP.

It is our argument that these 7 covariates help to explain why some sanitary jurisdictions make the most of SP, both in terms of affiliation and making new investments. Indeed, it is likely that the relative easiness with which it might be possible to expand the offer of health care services may be behind both the roll-out of the SP program and the response to it. A case in which a particular coverage pattern may have followed sanitary jurisdictions disposed to have also a particular response to the program. In other words, our basic assumption is that the infrastructure and personnel of the SSA that were already on the ground before the affiliation effort are both the major determinants of the SP roll out pace, and a clear sign of the difficulties each SJ face to translate financial means into basic resources needed to provide health care services.

In using the density of these resources as covariates in a PS framework, we are assum-

Figure 1.3: Causal Diagram Implicit in Observational Studies with Time-independent Confounding



Where $\mathbf{U}$ are unobserved exogenous variables, common causes of the level of exposure to treatment A and pretreatment covariates$\mathbf{X}$, and Y the outcome.

ing that, conditioning on them, the SP coverage can be considered as good as randomly assigned. In this way the treatment is therefore assumed independent of the potential outcomes, and so the treatment assignment mechanism is said to be ignorable. How to adequately control for these variables, on the other hand, depends on the causal assumptions behind the particular research design.

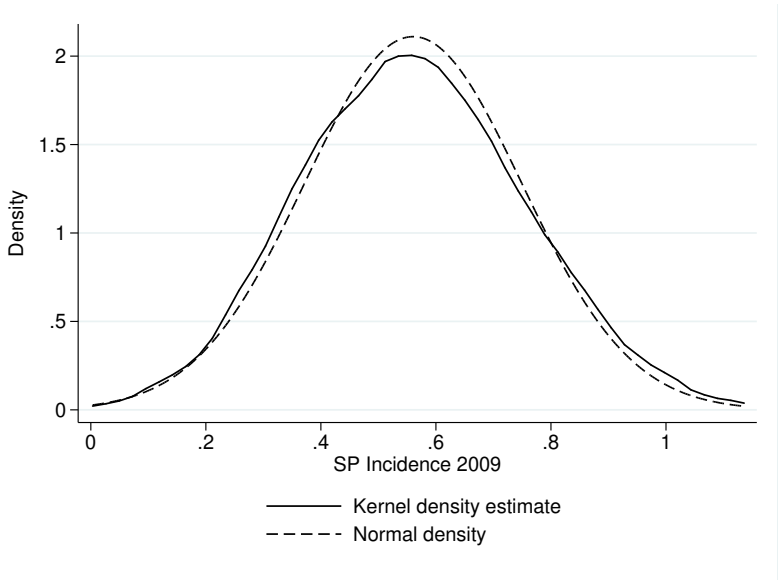
Propensity Score Methods with Time-independent Confounding

Figure 1.3 represents the usual (with simple treatment) causal assumptions behind the PS approach, where U represents unobserved (therefore connected by dashed arrows) exogenous variables, common cause to both the treatment and the outcome, as in our case investment performance or facility; A represents the causal variable, SP coverage; X a vector of covariates such as the density of doctors' offices, hospital beds, physicians and nurses; and Y some outcome variable, in our case the increment in the density of resources.

Coarsening an Underlying Treatment Variable

As for the classic PS approach, much of the work has focused on the case where units of analysis are exposed to one of two possible values of the causal variable, treatment

Figure 1.4: Distribution of SP Coverage Across Sanitary Jurisdictions in 2009



Source: Personal elaboration based on data from the administrative records of the Seguro Popular program

or control, at a given point in time, and values for an outcome are assessed some time subsequent to exposure [44, 45, 47]. To apply these traditional methods in evaluating SP, we need to operationalize SP coverage as a dichotomous variable for every SJ, much as Bosch and Campos [5] did for municipalities, while making the strong assumption of no-multiple-versions-of-the-treatment.[2]

Looking at figure 1.1 we can see that by 2009, SP had reached roughly 50% of the population without traditional health insurance, and figure 1.4 shows the distribution of this coverage across sanitary jurisdictions, exhibiting a fair number of observations on every treatment level in a relatively symmetric distribution around .5.

This suggest an almost natural way to dichotomize the SP coverage: around its median in 2009. In this way, the upper half of the distribution in 2009 is taken as the treatment

[2]Another potential approach for handling the multiple versions of SP coverage would be to focus on particular levels of exposure, however, our sample size for each version (no two are exactly the same) limits the effectiveness of this approach.

group, and the lower half as control.

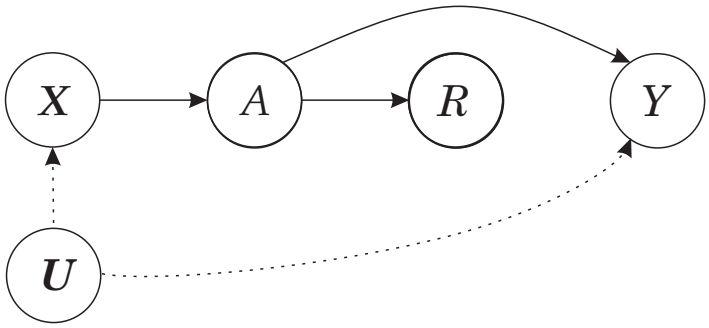
Using this dichotomized version of the SP coverage in 2009, we control for the 7 covariates corresponding to the year 2001, that is the year prior to the implementation of the SP program in its pilot phase.

Note that every estimate obtained following this design (Fig. 1.5), implicitly assumes that every SJ with an SP coverage roughly above 50% has being exposed to the same treatment regimen. Of course, compounding different levels of coverage into one single regimen misses the opportunity to exploit the most part of the SP coverage variability. Moreover, ignoring multiple versions of treatment can result in biased estimates due to inadequate control for confounding. Also, even if confounding is not a problem, this fuzziness in the treatment complicates the interpretation of causal quantities making it more subtle.

Nevertheless, even though we cannot say that the median dichotomized coverage corresponds unambiguously to a single homogeneous intervention, provided we have adequately controlled for confounding for the SP coverage (not just the median dichotomized version), we can potentially interpret the causal effect derived from this approach as the effect of comparing two randomized interventions. Think of a randomized control trial where levels of coverage are randomly assigned among SJs with the same characteristics, in one arm according to the distribution of levels above the median coverage, and, in the other arm, according to the distribution of levels under the median coverage. All of it under the (strong) assumption of no unmeasured confounding for the continuous SP coverage, something we can put to empirical test.

In this setting, the matching is performed by making use of two algorithms that automatize covariate balance optimization Genetic Matching (GM) [10], an evolutionary search algorithm to determine the weight each covariate is given; and Covariate Balancing Propensity Score (CBPS) [24], which models treatment assignment while optimizing the covariate balance in a Generalized Method of Moments (GMM) framework. The idea behind the CBPS is simple and consist of adding to the usual score vector of the treatment

Figure 1.5: Causal Diagram Implicit in Median Dichotomized Continuous Treatment



Where $\mathbf{U}$ are unobserved exogenous variables, common causes of the level of exposure to treatment A and pretreatment covariates$\mathbf{X}$, R the dichotomized (coarsened) version of the treatment (multiple values of A map onto a single value $R = r$), and Y the outcome.

assignment model, the difference in means between the treatment and control groups of every covariate as moment conditions. We also examine the limit case of balancing covariates dispensing altogether with the treatment assignment model. This non-parametric version of the CBPS methodology (NP-CBPS) can help us gain insight into what is to be expected from achieving a better balance.

Keep in mind that, in this case, the median dichotomized SP coverage acts as a mismeasured form of the original causal variable; that is, we have deliberately removed data in order to use traditional propensity score methods. However, there have been extensions to these methods in the last 15 years that allow us not to compromise data analysis in the face of recipients exposed to the same treatment in different degrees.

Generalized Propensity Score

Even though Hirano and Imbens [21] coined the term Generalized Propensity Score (GPS), several researchers proposed generalizations of the propensity score methodology for non-binary treatments before [43, 27, 26].

As in the case with the binary treatment, in estimating the GPS researchers model the distribution of the observed treatment assignment given pre-treatment covariates us-

ing a parametric model. In the practical implementation of this methodology, researchers often fit Gaussian distributions to continuous treatments like the SP coverage by means of Ordinary Least Squares (OLS). In other words, the GPS is equal to the treatment assignment model density, not necessarily Gaussian, evaluated at the observed treatment (t) exposure and covariates ($\mathbf{x}$), $f_{T|\mathbf{X}}(t, \mathbf{x})$.

There are several GPS methods to estimate the causal response of a continuous treatment [21, 12, 41] (for a comparison of the empirical performance of these methods see [52]). However, to allow for a better comparison with our dichotomous-treatment estimates, here we follow the GPS version of the Inverse Probability of Treatment Weights (IPTW) suggested by Robins, Hernán and Brumback [43]. Robins and colleagues pointed out that using $1/f_{T|\mathbf{X}}(t, \mathbf{x})$ as weights can lead to very unstable estimates, and suggested to use instead a more stable version of the weights: $W(t, x) = f_T(t)/f_{T|\mathbf{X}}(t, \mathbf{x})$, where the numerator corresponds to an estimate of the empirical distribution of the treatment. In particular, we use Robins' stabilized weights in regressing the outcome variables against SP coverage.

In estimating the stabilized weights, we have regressed SP coverage against our 7 covariates: doctors' offices, staffed and non-staffed hospital beds, physicians and nurses with and without day-to-day contact with patients, and all two-way interactions between these 7 variables, 35 variables in total, i.e.,

$$f_{T|\mathbf{x}} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[T - (\mathbf{X}'\boldsymbol{\beta})\right]^2\right\},$$

where β is the vector of regression coefficients, and $f_T(t)$ is estimated in the same fashion, only this time using an empty model.

Besides these OLS-based stabilized weights, we have also used a robust estimation procedure to estimate three different weights as proposed by Fong, Ratkovic, Hazalett and Imai [13], which is an extension of the CBPS methodology to continuous treatments. Like the CBPS for dichotomous treatments, this extension automatizes covariate balance optimization, which in this case signifies minimizing the correlation between the covari-

ates and the treatment. We estimate three different weights following this methodology [14]: two parametric models, one that gives equal importance to both correctly predict treatment-assignation and balancing covariates (CBPS-OVER), and a second one that privileges covariate balance over the probabilistic model (CBPS-EXACT); and the non-parametric approach that dispenses with models altogether only minimizing covariate balance (NP-CBPS) by means of Lagrange multipliers. It is important to note that none of the three CBPS-continuous weights use interactions, unlike the OLS stabilized weights, since our main concern is balancing only our 7 covariates.

Again, but this time using the original (continuous) SP coverage in 2009, we control for the 7 covariates corresponding to the year 2001. If covariates are indeed balanced across the different levels of SP coverage in the different CBPS pseudo-samples, we would expect that regressing each covariate against SP coverage results in a non-significant coefficient.

Heterogeneous Treatment Effects

Under ideal conditions, the Average Treatment Effect (ATE) provide a complete description of the relationship between the treatment and the outcome distribution across beneficiaries. However, it is well-known that averages may mask important subgroup differences, making it necessary to look into the conditions under which the same exposure to the treatment has a differential effect across segments of the population. This is of particular importance for social development programs attempting to reach those in greater need.

Quantile Treatment Effects

In exploring heterogeneous treatment effects, Quantile Treatment Effects (QTEs) play an essential role. QTEs give us the effect of the treatment, not on the mean, but on any given quantile of the outcome distribution. Thus, looking at the effect that a treatment has on different quantiles, researchers can go beyond the mean and assess the effect of the

treatment on the shape of the entire outcome distribution. Knowing whether the treatment has changed the skew in the outcome distribution or not, allows researchers to address questions about who are making the most of social programs, uncovering best practices and possible undesired effects on inequality.

A nice feature of the IPTW approach is that it can be easily modified for this end merely by resorting to a Quantile Regression [33]. The idea is basically the same as before only this time, instead of estimating the conditional mean by OLS, we have estimated several conditional quantiles of the outcome variables.

Interaction Effects

Also in the IPTW framework, investigating the interaction between the treatment and pre-treatment covariates can help in profiling recipients unable to benefit from the program as intended. The idea is to test, in a regression framework, whether the effect of SP coverage varies with the density of health care resources observed before SP. Where a positive, statistically significant coefficient estimate for the interaction between SP coverage and a pre-SP covariate, would suggest a stronger effect of SP as SJs are better off in terms of the covariate in question.

1.5 Results

Common Support

In order to compare only alike sanitary jurisdictions in everything but their SP coverage, following King and Zeng [29, 30] we use only the 194 SJs with the smallest mean distance to the rest of the observations in the data. In other words, we have dropped from the analysis the least comparable (farthest away in the covariate space) 17% of our SJs according to Gower's [16, 17] metric. This trimming results in a more compact (Fig. 1.6, Appendix Table A.1), less model dependent data set that allows better comparisons across

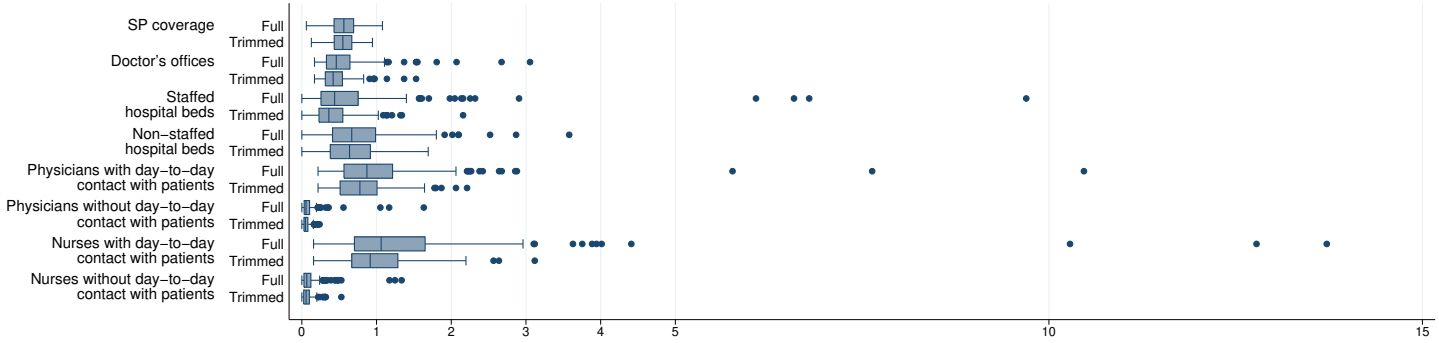
SJs.

Median Dichotomizing SP Coverage

Using the dichotomized version of the SP coverage in 2009, we match SJs on the 7 covariates corresponding to the year prior to the implementation of the SP program in its pilot phase: doctors' offices, staffed and non-staffed hospital beds, physicians with and without day-to-day contact with patients as well as nurses, also with and without day-to-day contact with patients.

From all three matching algorithms we can see we get a fairly balanced sample (Table 1.1) to estimate the effect of SP coverage. All differences in means are non-significant in the matched samples, reduction in bias is also quite sizable as well as the corresponding length of the equivalence regions. It is also worth noting that, after matching, none of the seven variables reject the hypotheses of equality of distribution. There are differences, however, in the variance of the variables across samples, where applying the GM algorithm results in groups more alike, especially regarding doctors' offices and nurses without day-to-day contact with patients, apparently at the expense of bias. All in all, these matching algorithms give us a range of balanced samples with different degrees of compromise between bias and variance.

Figure 1.6: Descriptive Statistics of the Sanitary Jurisdictions



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System).

The left and right side of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outliers as dots.

Other than SP coverage, variables are expressed per thousand population without traditional work-related health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

SP coverage refers to affiliates as proportion of the population without traditional health insurance.

The whole sample result from aggregating the municipality level data to the sanitary jurisdiction level.

The trimmed sample results from dropping the farthest away observations in the covariate space according to Gower's metric [16, 17]. The Gower distance between two SJs, is the average of the absolute differences between the covariates, as a proportion of the range of the corresponding covariate in the data.

Table 1.1: Covariate-Balance for Dichotomous Treatment on Sanitary Jurisdictions[a]

Variable		Mean			% red	Diff. in means[d]		Eq. region[e]		K-S test	V(T)/
(2001)[b]	Sample[c]	Treated	Control	% bias	bias	t	$p > t$	diff	std diff	p-value[f]	V(C)[g]
DO	UM	0.50	0.40	52.20		3.63	0.00	0.16	80.48	0.00	1.18
	GM	0.45	0.44	5.40	89.70	0.38	0.71	0.07	33.49	0.56	0.90
	CBPS	0.46	0.43	15.30	70.60	0.99	0.33	0.09	45.92	0.11	0.67*
	NPCBPS	0.45	0.45	0.10	99.80	0.01	0.99	0.07	33.23	0.23	0.53*
SHB	UM	0.40	0.45	-16.90		-1.18	0.24	0.14	45.26	0.75	0.57*
	GM	0.42	0.44	-6.10	64.00	-0.46	0.65	0.10	32.39	0.39	0.85
	CBPS	0.42	0.42	-1.90	88.80	-0.14	0.89	0.09	28.99	0.72	0.81
	NPCBPS	0.43	0.42	0.20	98.60	0.02	0.99	0.08	26.88	0.51	0.91
NSHB	UM	0.81	0.54	78.00		5.43	0.00	0.36	106.33	0.00	1.19
	GM	0.68	0.66	7.20	90.80	0.51	0.61	0.12	35.11	0.75	1.07
	CBPS	0.71	0.66	14.60	81.30	0.94	0.35	0.15	45.25	0.51	0.79
	NPCBPS	0.67	0.67	-0.10	99.80	-0.01	0.99	0.11	30.80	0.67	0.69

Table continues

Variable		Mean			% red	Diff. in means[d]		Eq. region[e]		K-S test	V(T)/
(2001)[b]	Sample[c]	Treated	Control	% bias	bias	t	$p > t$	diff	std diff	p-value[f]	V(C)[g]
PWC	UM	0.87	0.75	34.60		2.41	0.02	0.22	62.93	0.01	1.00
	GM	0.80	0.79	2.50	92.80	0.18	0.86	0.10	29.37	0.91	1.03
	CBPS	0.83	0.78	13.00	62.30	0.93	0.35	0.14	40.75	0.32	0.90
	NPCBPS	0.81	0.81	0.10	99.70	0.01	0.99	0.10	28.81	0.76	0.81
PWOC	UM	0.06	0.06	-13.70		-0.95	0.34	0.02	41.98	0.62	0.60*
	GM	0.06	0.06	-5.10	62.70	-0.38	0.71	0.02	31.59	0.81	0.88
	CBPS	0.06	0.06	-0.60	95.60	-0.04	0.97	0.01	28.76	0.16	0.68
	NPCBPS	0.06	0.06	0.20	98.60	0.01	0.99	0.01	29.36	0.14	0.73
NWC	UM	1.07	0.97	20.40		1.42	0.16	0.24	48.69	0.02	0.77
	GM	1.02	1.02	1.20	94.20	0.08	0.93	0.14	28.89	0.56	0.93
	CBPS	1.05	0.99	12.30	39.80	0.85	0.39	0.20	40.58	0.45	0.85
	NPCBPS	1.02	1.02	0.20	99.00	0.01	0.99	0.14	29.02	0.90	0.76

Table continues

Variable		Mean			% red	Diff. in means[d]		Eq. region[e]		K-S test	V(T)/
(2001)[b]	Sample[c]	Treated	Control	% bias	bias	t	$p > t$	diff	std diff	p-value[f]	V(C)[g]
NWOC	UM	0.06	0.08	-23.70		-1.65	0.10	0.04	52.00	0.21	0.50*
	GM	0.06	0.07	-8.10	66.00	-0.63	0.53	0.02	33.14	0.50	0.79
	CBPS	0.07	0.07	-8.20	65.30	-0.62	0.54	0.02	34.48	0.79	0.66*
	NPCBPS	0.07	0.07	0.10	99.60	0.01	0.99	0.02	27.56	0.76	0.79

Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

* if variance ratio outside [0.67; 1.50] for UM and [0.67; 1.50] for all matching algorithms

[a] Treatment variable results from dichotomizing Seguro Popular coverage in 2009 around its median.

[b] Per thousand population without traditional (work-related) health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy), where DO stands for doctors' offices, SHB for staffed hospital beds, NSHB for non-staffed hospital beds, PWC for physicians in day-to-day contact with patients, PWOC for physicians without day-to-day contact with patients, NWC for nurses in day-to-day contact with patients and NWOC for nurses without day-to-day contact with patients.

[c] UM sample refers to the 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, GM to the sample matched using Genetic Matching, CBPS to the sample matched using Covariate Balance Propensity Score and NPCBPS to the sample matched using Non-parametric Covariate Balance Propensity Score.

[d] As suggested by t-statistic, none of the tests in the table take into account the statistical error associated to the matching procedure.

[e] Equivalence region as the largest difference at which the null hypothesis of difference-in-means would be rejected with a confidence of 95%, both in units of the variable tested and standardized by the average variance of treated and no-treated. Unlike the difference-in-means test (left columns), where the equivalence is the null and difference the alternative, here we only reject the null if sufficient data exists to demonstrate otherwise. The region does not take into account the matching algorithm.

[f] Kolmogorov-Smirnov equality-of-distributions test, the corrected combined p-value is reported.

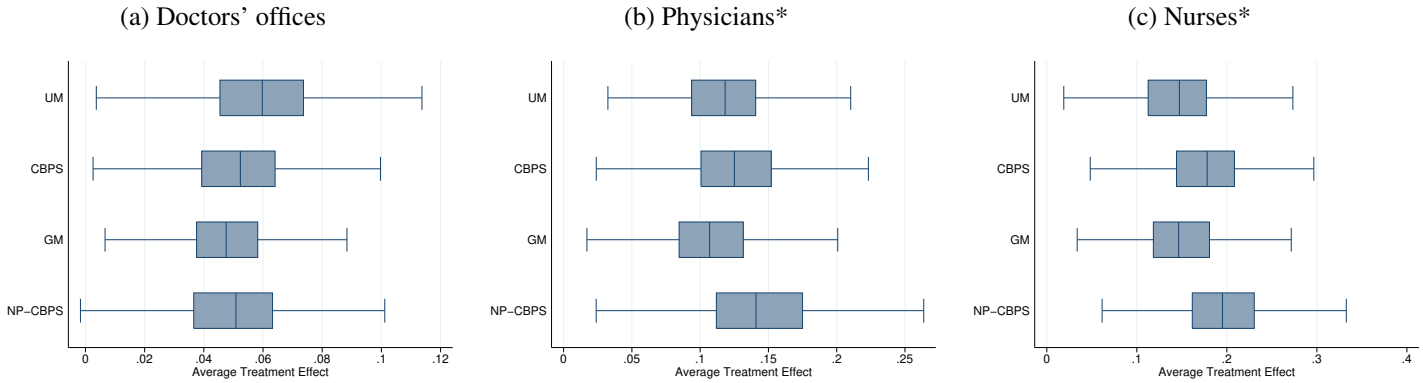
[g] Variance ratio of treated over non-treated. An asterisk is displayed for variables that have variance ratios that exceed the 2.5th and 97.5th percentiles of the F-distribution with number of matched treated minus 1 degrees of freedom.

Looking at the change in the density of doctors' offices, physicians and nurses from 2001 to 2010 we can see (Fig. 1.7, Appendix Table A.2) our estimates suggest that SP has had a clear positive impact on SSA's resources.

However, several important points concerning the interpretation of these results merit attention. First, we would be wrong in assuming these estimates correspond to the average treatment effect of SP *full* coverage, for the information lost in dichotomizing the treatment has led us to identify a different causal quantity difficult to pinpoint [19, 51], thus limiting data analysis. Secondly, by definition, the causal effect of the median dichotomized coverage is null (Note that in Fig. 1.5 there is no arrow from R to Y since, after all, we only dichotomized the causal variable). What the results of this approach are quantifying is the association between the dichotomized coverage and the outcome variables, mediated through their common cause: the original continuous SP coverage. As noted in Section 1.4, these effect estimates can be interpreted as an estimate of what would have been observed in the randomized trial in which, within strata of covariates $\mathbf{X} = \mathbf{x}$, an SP coverage is randomly assigned to every SJ from the observed distribution of "levels of coverage" in the population among those with coverage, in one arm superior to the median, and in the other arm less than the median. But this interpretation is predicated on the assumption that we have adequately controlled for confounding for the underlying continuous version of the treatment variable; that is, the assumption of no unmeasured confounding for SP coverage.

In this regard, it is important to note that balancing covariates with respect to the median dichotomized SP coverage does not necessarily achieve "balance" across different values of the original continuous causal variable. We can see (Figs. 1.8 and 1.8, Appendix Table A.3) how this is not the case for doctors' offices, non-staffed hospital beds, physicians and nurses in day-to-day contact with patients, where, despite the fact that in the pseudo-samples generated with every matching algorithm, the correlation between the dichotomized SP coverage and these covariates is statistically indistinguishable from zero, the correlation still remains when we look at the original (continuous) SP coverage. This

Figure 1.7: Average Treatment Effects of Dichotomized SP coverage on Sanitary Jurisdictions by Matching Algorithm[ab]



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

The left and right side of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

SP coverage refers to the number of affiliates in 2009 as proportion of the population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

* In day-to-day contact with patients

[a]Effects on the 2010-2001 increment in the density of resources, per thousand population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy. Where the treatment variable results from dichotomizing Seguro Popular coverage in 2009 around its median.

[b]UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, CBPS to the sample matched using Covariate Balance Propensity Score, GM to the sample matched using Genetic Matching and NP-CBPS to the sample matched using Non-parametric Covariate Balance Propensity Score, all with respect to dichotomized treatment.

is so even in the NP-CBPS pseudo-sample, where the correlation with the dichotomized treatment is driven to virtually zero.

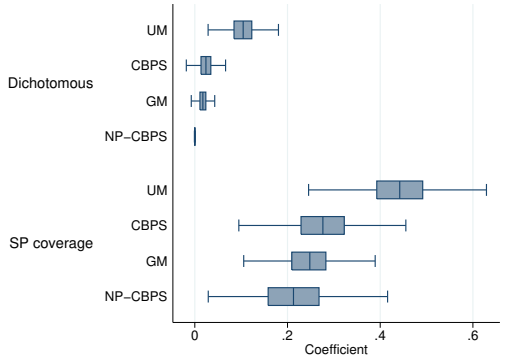
In other words, the data suggests that it would be a mistake to assume that we can treat SP coverage as randomly assigned within our treatment groups in the matched samples, seriously undermining any causal interpretation of these estimates related to the original SP coverage. This challenges arise because the causal effect of aggregated/compound treatments (in our case dichotomized) depends on the distribution of the original treatment in the population [19], which we have not taken into account (as if utterly random with equal probability). We could still recover the causal interpretation of our estimates if we were to resort to other statistical methods that allow us to consider the distribution of "coverage levels" given the dichotomized coverage and covariates, that is, twice the conditional probability density of the coverage on each side of the median; but this would bring up the question of why not estimate the causal effect of the original SP coverage itself instead.

This leaves us with a poorly defined intervention for the purposes of thinking about causal effects (we do not know which causal effect is being estimated, if any), let alone actionable policy recommendations. However, these estimates still render statements such as "being in the upper half of the distribution of levels of coverage is on average better than being on the lower half", provided that these statements are understood with reference to the current, unknown (otherwise we would have taken them into account) policies for assigning levels of coverage. On the other hand, if one does not know what exactly these policies consist of, then one does not know what kind of intervention will have an effect that bears much resemblance to the estimate just obtained, making it meaningless for public policy.

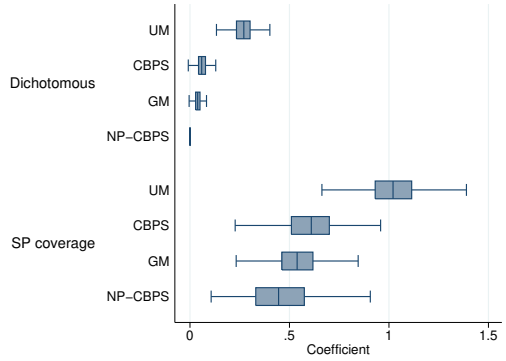
If we are not to compromise possible insights discarding valuable data in aggregating coverage levels, we need to resort to other propensity score methods, different from those designed for binary casual variables.

Figure 1.8: Covariate-Balance for SP coverage with Balanced Samples with respect to Dichotomized Treatment on Sanitary Jurisdictions

(a) Doctors' offices

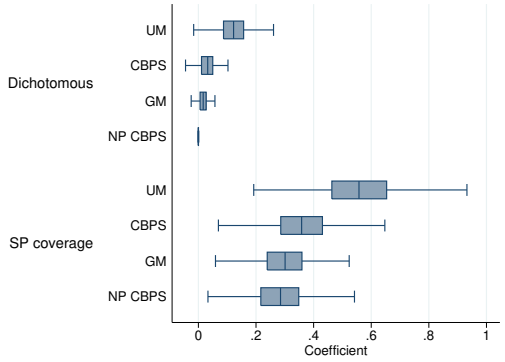


(b) Non-staffed hospital beds

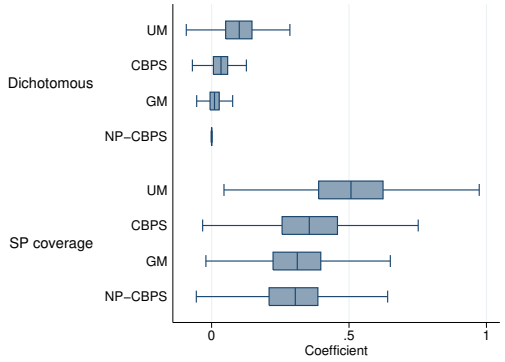


(*Figure continue*s)

(c) Physicians*



(d) Nurses*



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records.

The left and right side of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Coefficients result from linear regressions of each covariate in 2001 on the dichotomized and original SP coverage in 2009, weighted accordingly to the pseudo-sample in question.

Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

* In day-to-day contact with patients

SP coverage refers to affiliates in 2009 as proportion of the population without traditional health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy). Dichotomous treatment variable results from dichotomizing Seguro Popular coverage in 2009 around its median.

UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, CBPS to the sample matched using Covariate Balance Propensity Score, GM to the sample matched using Genetic Matching and NP-CBPS to the sample matched using Non-parametric Covariate Balance Propensity Score, all with respect to dichotomized treatment.

Continuous Treatments

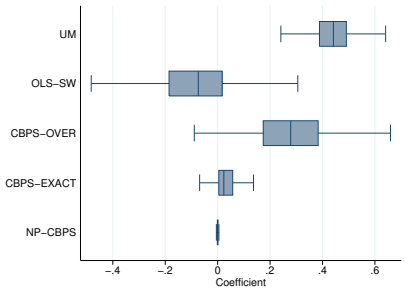
Again limiting our estimates to the trimmed 194 SJs sample, using the original continuous version of the SP coverage in 2009, we can see (Fig. 1.9, Appendix Table A.4) that, unlike the pseudo-samples generated using the PS methodology for dichotomous treatments, the pseudo-samples generated using the several weighting algorithms for continuous treatments (as described in Section 1.4), indeed break the association observed between the covariates and the SP coverage, although the CBPS-OVER algorithm leaves room for doubt at the 90% confidence level. [3]

Looking again at the change in the density of doctors' offices, physicians and nurses from 2001 to 2010, this time using the pseudo-samples generated for the original SP coverage, we see (Fig 3.3, Table A.5 in appendix) that these new estimates also suggest that SP has had a positive impact on SSA's resources (all estimates statistically different from zero), and unlike the dichotomous estimates, we can interpret them as the Average Treatment Effect of full SP coverage on the SJs. Attending to the smallest numbers, which also correspond to the best balanced sample, our estimates suggest that on the average, for our trimmed sample, full coverage of SP translates into roughly 2 doctors' offices, 6.5 physicians and 8 nurses, both in day-to-day contact with patients, per ten thousand population. These numbers may seem small, but keep in mind that in 2001, in these same 194 SJs, the average density of these same resources was 4.5, 8 and 10 respectively. Also of importance is to note that these figures differ substantially from the "naive" OLS estimates that overestimate the impact of SP on doctors' offices by 26%, and underestimate its impact on physicians and nurses in day-to-day contact with patients by 23% and 34% respectively. Let's not lose sight of the fact that these estimates belong to the average, that is, not to a SJ in particular, and that some SJs may be doing better than others in expanding these resources for different reasons.

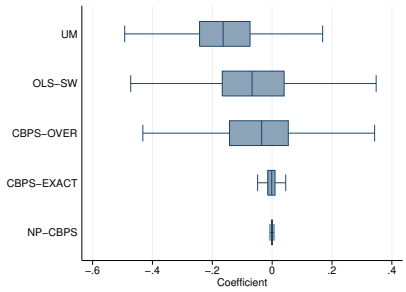
[3] We see this regressing each covariate against SP coverage before and after appropriately weighting the trimmed sample.

Figure 1.9: Pre-Treatment-Covariate-Balance for SP coverage on Sanitary Jurisdictions by Weighting Algorithm

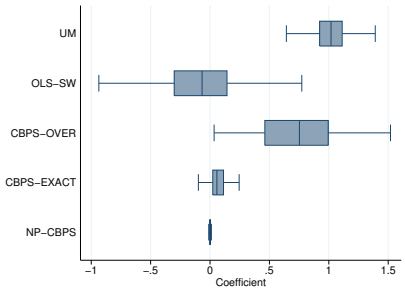
(a) Doctors' offices



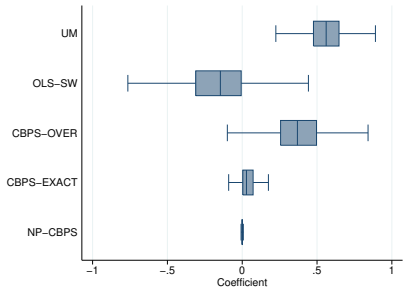
(b) Staffed hospital beds



(c) Non-staffed hospital beds

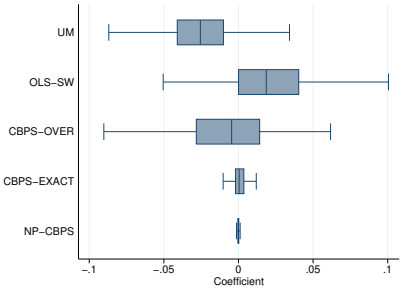


(d) Physicians*

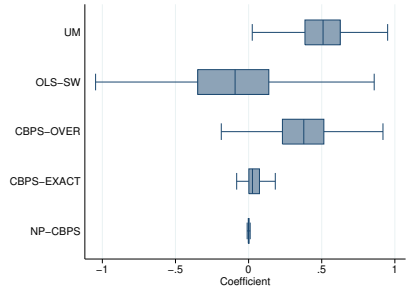


(Figure continues)

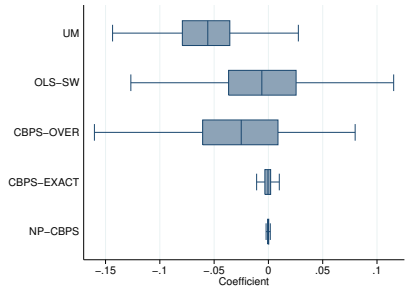
(e) Physicians**



(f) Nurses*



(g) Nurses**



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

The left and right side of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Coefficients result from linear regressions of each covariate in 2001 on the original SP coverage in 2009, weighted accordingly to the pseudo-sample in question.

Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

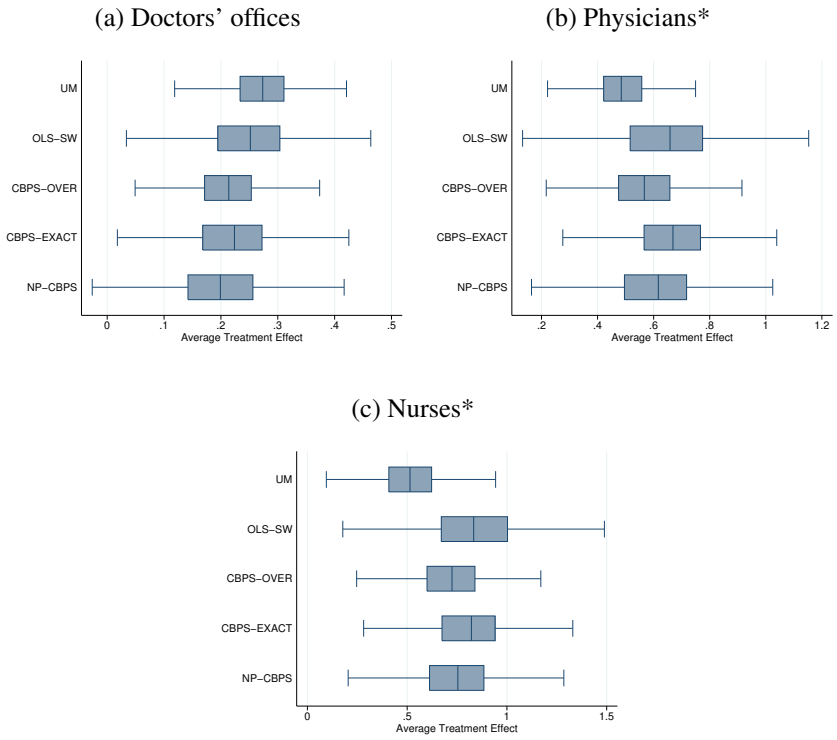
* In day-to-day contact with patients

** Without day-to-day contact with patients

SP coverage refers to Seguro Popular affiliates in 2009 as proportion of the population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy

Where UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance using Lagrange multipliers.

Figure 1.10: Average Treatment Effects of SP coverage on Sanitary Jurisdictions by Balancing Algorithm



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

The left and right side of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Estimates result from linear regressions of each outcome variable on the original SP coverage in 2009, weighted accordingly to the pseudo-sample in question.

Effects on the 2010-2001 increment in the density of resources, per thousand population without traditional health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

* In day-to-day contact with patients

SP coverage refers to affiliates in 2009 as proportion of the population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

Where UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance using Lagrange multipliers.

Heterogeneous Treatment Effects

As noted in Section 1.4, IPTW methods can be easily modified to go beyond the Average and into the Quantile and Interaction Treatment Effects to profile segments of the population with heterogeneous responses to SP full coverage. Looking at the QTE, we can see from our estimates (Fig. 1.11, Table A.6 in appendix) that the effect of full SP coverage is smaller in the lower part of the outcome distributions. This suggests that SP has had more than a central location effect on the outcome distributions, as would correspond to similar magnitudes across quantiles. On the contrary, treatment effects are quite heterogeneous along the distribution of the outcome variables, being clearly more concentrated as we move towards the upper quantiles in the case of doctors' offices and the third quartile for physicians and nurses. That is to say that the QTEs show that SP program has widened the spread of the outcome distributions, exacerbating the increment in the density of health workforce and infrastructure as SJs were better off making progress on this front. This is particularly so for doctors' offices, where the effect of SP is statistically indistinguishable from zero for the first decile, and grows steadily up to twice the mean for the last one. It is important to note that physicians and nurses also exhibit the smallest coefficients in the first two deciles of their respective distributions, being mostly non-significant for the first decile, with a third quartile roughly 1.5 times the average.

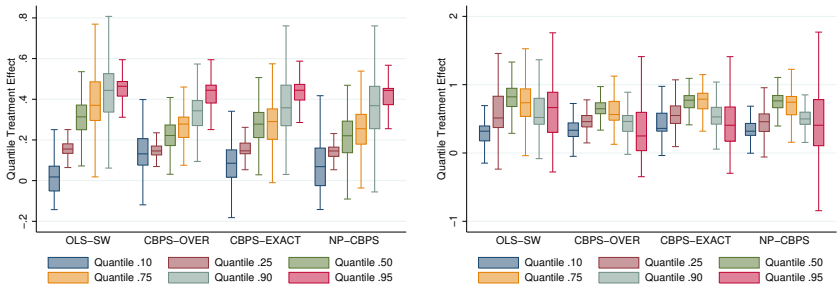
It is important to remember, before jumping to conclusions, that the outcome variables examined correspond to changes in the density of SSA's resources, not the density of the resources themselves. Whether SP has contributed or not to a more unequal distribution of health care resources in Mexico depends on which are those SJs reaping the most benefits from program. If the SJs benefiting the least from the program were precisely those less advantaged in terms of health care resources, then the SP program would likely have contributed to the inequality in the distribution of these resources. If, on the other hand, the least advantaged jurisdictions were making the most of the program, it would have contributed to alleviate this same inequality.

In this regard, our interaction estimates (Fig. 1.12, Table A.7 in appendix) between SP

Figure 1.11: Quantile Treatment Effects of SP coverage on Sanitary Jurisdictions by Balancing Algorithm

(a) Doctors' offices

(b) Physicians*



(c) Nurses*

Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

The bottom and top of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Estimates result from linear quantile regressions of each outcome variable on the original SP coverage in 2009, weighted accordingly to the pseudo-sample in question.

Effects on the 2010-2001 increment in the density of resources, per thousand population without traditional health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

* In day-to-day contact with patients

SP coverage refers to affiliates in 2009 as proportion of the population without traditional (work related) health insurance.

Where OLS-SW corresponds to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance using Lagrange multipliers.

coverage and the baseline levels of the resources suggest that SP has had stronger effects boosting the investment on physicians and nurses in day-to-day contact with patients as the density of physicians without day-to-day contact with patients in 2001 grows larger in the SJs. Bear in mind that, in 2001, more than half of the physicians without day-to-day contact with patients were engaged in administrative tasks, with the rest dispersed among educational activities, epidemiologists and anatomo-pathologist. That is to say that the number of physicians without day-to-day contact with patients is closely related to institutional development and higher-level health care services. In 2001, 64% of the physicians without day-to-day contact with patients worked at Specialized (Inpatient) Hospitals.

Looking closely at the distribution (1000 repetitions) of our estimates for physicians and nurses in Figure 1.12, we can see that SP seems to have a higher effect on SJs with lesser densities of doctors' offices, which would indicate that poorer SJs favor the investment in health workers providing clinical care.

Taken together, the heterogeneity in SP treatment effects would suggest that the SJs making the most of SP are precisely those better off in terms of the material and human resources examined. These results seem to evidence the apparent incapability of SP to work against the inequality in the distribution of resources associated with to the provision for healthcare services.

1.6 Conclusions and Discussion

First of all, our results point in the same general direction of those of Bosch and Campos [5]: on the average, the SP program has translated to increases in the resources allocated to provide health care. However, a number of the precise technical details differ. In this paper we've taken a step further in various directions: on the one hand, in assessing the impact of SP on the infrastructure and human resources of the SSA, we have used the basic regional administrative unit in charge of the operation of healthcare services and programs as units of analysis: the sanitary jurisdiction. On the other hand, we have also

used the coverage of Seguro Popular as a continuous causal variable, which has allowed us to gain insight into the effects of SP otherwise impossible.

We have shown evidence that average results hide a great deal of information regarding the effects of SP. Through a heterogeneity analysis we have shown that quite possibly SP has had distributional effects on the resources involved in the provision for healthcare services. Our results suggest that the sanitary jurisdictions that were better off in terms of these resources before SP, are precisely those making the most of the program. The fact that SP may be leaving behind the most vulnerable geographic areas in the country is of major concern from a public policy perspective, for it points to the need for complementary public health policies.

We have also shown strengths and weakness of different research designs according to their causal and statistical assumptions, we believe it is for the benefit of applied researchers to be reminded of how different versions of treatment may affect (bias) inferences and their interpretation, a problem generally not possible to eliminate entirely in making causal inference, against which we must remain always vigilant. We cannot overemphasize the importance of this when informing public policy.

In this regard, it is useful to be explicit about the statistical and causal assumptions behind all of our estimates. Most importantly, our treatment variable, SP coverage, makes no distinction between different treatment histories. As long as two sanitary jurisdiction exhibit the same program incidence in 2009, our approach treats them as if they had the same amount of treatment exposure, which may not be true in a very direct sense. It may well be that even though two municipalities show the same coverage in 2009, say 80 percent of the population without traditional (work-related) health insurance, one of them has just recently achieved this while the other has been showing this level of coverage for several years. In this context it may seem somewhat inappropriate to consider that these two territories have been exposed to the same benefits of the program. Also, overlooking treatment histories altogether has led us in turn to omit considering their possible interaction with confounders at different points in time; that is, in all of our estimates we have ignored the

possibility that SP coverage is time-varying.

Our analysis focused on contrasting different assumptions help us to consider, by analogy, the implications of overlooking time-dependence, as this can be seen as another way of compounding (aggregating) treatments that may lead to inadequate control for confounding.

To see why this is so, consider that conditioning only on covariates from 2001, the year prior to the implementation of the SP program in its pilot phase, we have implicitly assumed SP coverage as a fixed, time-independent (non-dynamic) treatment, and that the covariates from 2002 to 2008 do not provide relevant information for the causal analysis at hand (Fig. 1.13).

However, let's keep in mind that, even in 2009, the affiliation effort of SP had reached over 31 million people, of which 28 million lived in our trimmed sample of 194 SJs. Naturally, affiliating this many people took years, and it is quite likely that during those 7 years, the SP's roll-out strategy reacted to the changing conditions SP itself helped bring about.

If this had actually happened, suppose that in fact $\mathbf{X}_t$ lies in the causal path (is an effect of) exposure A_{t-1}as depicted in Fig. 1.13, there would be some uncontrolled confounding in our design running through the dashed lines. However, note that covariates from 2002 to 2008 are no longer strictly pre-SP covariates with respect to the coverage attained by 2009, only with respect to the coverage attained after the specific year we are looking at. Indeed, covariates from 2006, for example, are pre-treatment with respect to the increment in coverage observed after that year, but at the same time are post-treatment with respect the coverage observed already by that same year. In that sense, all the covariates from 2002 to 2008 are, at the same time, pre-treatment and post-treatment with respect to some part of the coverage observed in 2009. Thus, we would be mistaken in including them in our estimation of the GPS, for we would be severing one possible channel through which the treatment may be affecting the outcome (the dotted lines in Fig. 1.13), a clear example of post-treatment bias [4]. On the other hand, by not including them, we are probably

biasing our estimates of the impact of SP by leaving out relevant confounders from the analysis.

However, behind our assumed homogeneity in SP coverage is the current state of methodology on GPS methods, which have yet to reach the case of continuous time-depending treatments.[4] There are further generalizations of the GPS methodology that exploits precisely the time-series cross sectional nature of data sets [4, 43, 14], but so far they've focused only on time-varying dichotomous treatments. In a first impulse, researchers may try to dichotomize SP coverage every year in order to apply these methodologies, but this strategy leads to missing the opportunity to use all the available information just like in the fixed case. Consider the case where the particular level of coverage at time t may be a strong confounder for the dichotomized coverage at subsequent times. Plus, there is no clear way to dichotomize SP coverage following the same rule every year, difficulting even more the interpretation of causal quantities.

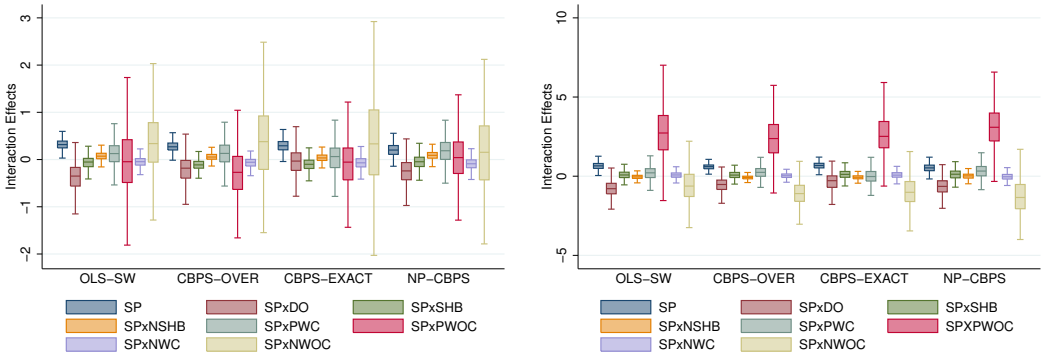
Let this discussion serve as motivation to keep pushing forward Propensity Score methods, a task we take upon ourselves in the next chapters.

[4] Causal inference for time-varying continuous treatments requires estimating the conditional probability density of continuous treatment histories.

Figure 1.12: Interaction Effects of SP coverage with Pre-Treatment Covariates on Sanitary Jurisdictions by Balancing Algorithm

(a) Doctors' offices

(b) Physicians*



(Figure continues)

(c) Nurses*

Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

The bottom and top of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Estimates result from linear regressions of each outcome variable on the original SP coverage in 2009 and its interaction with all 7 per-treatment (2001) covariates, weighted accordingly to the pseudo-sample in question, where SP stands for SP coverage, SPxDO for the interaction between SP coverage and doctors' offices, SPxSHB for the interaction between SP coverage and staffed hospital beds, SPxNSHB for the interaction between SP coverage and non-staffed hospital beds, SPxPWC for the interaction between SP coverage and physicians in day-to-day contact with patients, SPxPWOC for the interaction between SP coverage and physicians without day-to-day contact with patients, SPxNWC for the interaction between SP coverage and nurses in day-to-day contact with patients and SPxNWOC for the interaction between SP coverage and nurses without day-to-day contact with patients.

Effects on the 2010-2001 increment in the density of resources, per thousand population without traditional health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

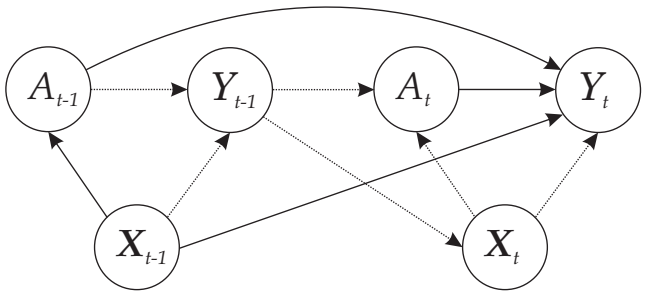
Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

* In day-to-day contact with patients

SP coverage refers to affiliates in 2009 as proportion of the population without traditional (work related) health insurance.

Where UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance by means of Lagrange multipliers.

Figure 1.13: Causal Diagram Illustrating Time-dependence



Where X are time-varying confounders, A the exposure history to treatment and Y the outcome at every point in time. Dashed lines represent the causal assumption of no causal effects implicit in using only pre-SP covariates in estimating the causal effect of A.

Chapter 2

Time-Varying Continuous Interventions: Seguro Popular and the Supply of Health Care Services

This Chapter presents a new semiparametric procedure to analyze time-varying continuous interventions. This is accomplished by bringing together the literature on continuous and dynamic treatments. Our approach allows the researcher to estimate Mean and Quantile Dose-Response Functions by applying local regression methods to appropriately weighted samples that control for time-dependent confounding. As an empirical application of the proposed method, we analyze the effects of the Mexican universal health insurance program, Seguro Popular, on key variables associated with the provision of healthcare services by the Health Ministry (Secretaría de Salud, SSA). We find compelling quantitative evidence that the program has proven most helpful in less vulnerable territories, leaving behind those in greater need.

2.1 Introduction

In assessing whether or not a program has achieved its intended results, the impact evaluation literature has heavily relied on causal inference methods focused on the effect of a single homogeneous action at a single point in time. Observing units (households, people, territories) at only one or two points in time, researches are typically left to assume that the program under scrutiny –the action that in the analysis is to be interpreted as cause– took place in a "single-shot" and without discernible variation among recipients. Even though implicit most of the time, this is a crucial assumption. The farther away we are from meeting this assumption and consider programs that unfold over time with varied intensity, the more biased conventional analytic methods will be and ultimately result in estimates that may fail to have a causal interpretation.

Naturally, implementing a program always takes time. When a program evolves over time in a deterministic way, assuming it as a single-shot intervention does not represent a problem in itself. If, on the other hand, the effectiveness of the intervention at earlier stages of implementation were to affect its nature at subsequent stages[1], assuming the program as a single-shot intervention could lead to comparisons between incomparable units. The problem arises because, in a non-deterministic time-varying intervention, the dynamic of the intervention itself generates potential confounders, which are impossible to control for before the program takes place. It is important to note that even though in a time-varying intervention like the one described before, units may end up with different "exposures" to the program –as different intervention histories are observed–, this is not the reason why it is impossible to control for variables that could bias our causal estimates using only baseline covariates. If the intensity or duration of an intervention is somehow determined for every unit at the outset of the program, it may be possible to control for confounders

[1]There is nothing abnormal in this situation, nearly all treatments of epidemiological interest present these characteristics –imagine a drug whose dose is readjusted according to the patient's clinical response. Recently these tools, indeed originally developed in biostatistics and epidemiology, have been applied in political science [4] to study the effectiveness of a candidate's decision to "go negative" in political campaigns, an inherently dynamic process.

using only preprogram covariates, as the program would still conform to the assumption of a single-shot intervention (see for example the methods discussed in [52]).

To properly assess the causal effect of a dynamic process in an observational study, researchers need to collect repeated measurements of the same units at several points in time. When the data collected in this fashion provides information of all covariates with the potential to both affect the future exposure and be affected by previous exposure to the program at a given point in time –time-varying confounders–, researchers are in position to estimate the effect of such dynamic interventions.

To this end, Robins, Hernán and Brumback[43], RHB from now on, presented a set of tools to estimate the causal effects of dynamic processes from Time-series Cross-Sectional (TSCS) data. These tools explicitly model the dynamic selection inherent in these time-varying processes, overcoming the biases derived from wrongly assuming a single-shot kind of intervention. RHB's methodology applies Inverse Probability of Treatment Weights (IPTW) to a class of semiparametric models called Marginal Structural Model (MSM). In these MSM, each unit is weighted by the inverse probability of its observed intervention history, creating a pseudo sample where dynamic selection is eliminated. Then, using this pseudo sample, the MSM dictates the form of the relationship between the histories and the outcome of interest.[2]

However, typically in the MSM literature, the intervention histories are coded as sequences of a binary variable; that is, histories of homogeneous (of uniform nature) interventions that either take place or not at every point in the time sequence [7], thus leaving out the case of interventions that, at every stage, may occur with varied intensity –best coded as a continuous variables.

Here we propose to fill this gap by being the first to take the applications of RHB's methodology from binary to continuous treatments, and then use the resulting IPTW to estimate Mean and Quantile Dose-Response Functions (DRFs) in these dynamic contexts.

[2]Keep in mind that, with time-varying continuous interventions, estimating the causal effect of every intervention history proves an impossible task in most cases as their number grows geometrically with the number of stages.

As an empirical application, we use this procedure to estimate the effect of the Mexican universal health insurance program, Seguro Popular (SP), on key variables associated with the provision of healthcare services by the Secretaría de Salud –Mexican Ministry of Health– (SSA).

In line with previous research [5], we found that the program has had a positive effect on the infrastructure and human resources of the SSA. However, the methodology implemented here allowed us to address a richer set of causal questions as we are now in position to analyze in detail the response to the gradual rollout of the program. As a result of this expanded scope we also found evidence that the program has had heterogeneous effects across the regional administrative units of the SSA, exhibiting greater impacts on those units with greater density of human and material resources, thus leaving behind those units with scarcer resources.

The paper proceeds as follows. Section 2.2 describes how RHB's methodology extends the methods focused on the effect of a single homogeneous action at a single point in time. Section 2.3 introduces the estimation procedures behind this methodology. Section 2.4 describes the weighting approach to estimating dynamic mean and quantile dose-response functions, and section 2.5 applies the techniques for estimating the effect of the SP program on the supply of healthcare services. Section 3.6 concludes with some public policy implications of the analysis.

2.2 Causal Inference with Time-Varying Continuous Interventions

In assessing whether or not a program has achieved its intended result, a sobering well known fact is that association is not causation. However, in randomized evaluations, association measures can be interpreted as effect measures because randomization ensures that the units are interchangeable whatever their exposure to the program. On the other

hand, in nonrandomized evaluations, association measures cannot be interpreted as effect measures because units with different exposures to the program are not generally interchangeable. In order to establish to what extent a particular program—and that program alone—contributed to the change in an outcome, after the program has been implemented, we must turn to methods that help us imitate a randomized evaluation in generating a valid (interchangeable) group of comparison.

In overcoming this challenge, impact evaluation methods have heavily relied on assuming the program under analysis as a single homogeneous intervention. This is a suitable framework for many programs whose intervention is reasonably standardized and takes place only once at fairly the same time for every recipient of the program. There are, however, many programs that don't conform to this assumption. Programs that evolve over long periods of time with different intensity for different units at every stage of their implementation, reacting to the changes in the environment that themselves cause to take place. The existence of the later kind of interventions has motivated the literature on causal effects of time-varying exposures.

To separate the causal effect of a dynamic program from mere association, we must, as always, be confident that the observed correlation is not due to some other variable. This risk usually referred to in the literature as confounding, is not different from single homogeneous interventions. However, in dynamic settings confounders may also share the time-varying nature of the intervention. A time-varying confounder can both affect future exposure to the program and be affected by past exposure, and impact evaluation methods that in this case wrongly assume single homogeneous interventions provide no way of removing the bias due to this kind of confounders [4], no matter the kind of data available.

By a generalization of IPTW estimators [22] for longitudinal studies, RHB presented a framework to estimate the effects of dynamic interventions (dynamic causal inference) in the presence of time-varying confounders from TSCS data. Like other impact evaluation methods, RHB's approach relies on the ability to identify and, most importantly, measure

all possible confounders, in this case repeatedly over time; that is, on a dynamic version of the well-known "selection on observables" or "ignorability" assumption. Indeed, unbiased estimation of causal effects by IPTW is impossible in the presence of unmeasured confounding factors at any stage of the implementation of the program. Even more, if some of the confounders are unobserved, the causal effect of a program may not be identifiable, in which case there is no procedure that can estimate it consistently. Needless to say we can never be sure that we observe all potential confounders. As is usual in empirical research, all we can do is turn to known theory and prior evidence to inform us of possible confounders.

In the counterfactual framework of causal inference, this assumption, also referred to as sequential ignorability [4] or conditional interchangeability [42], states that the exposure level at stage t is statistically independent (ignorable) of the potential outcomes, conditional on the covariate and exposure histories up to that point.

To express this assumption in its usual statistical form, let i index recipients of the program, with $i = 1, \ldots, n$ and t the stage of the intervention, taking possible values $1, \ldots, T$, where T is the last stage of the program recorded in the data. At every stage t of the program, recipients are observed receiving the exposure level a_{it} to the benefits of the program; that is, one possible realization of the exposure variable A_t. Collecting all of the observed exposures to the program for a given recipient up to stage t of the intervention gives us the history of exposure, $\underline{a}_{it} = (a_{i1}, \ldots, a_{it})$, an instance of the collection of variables $\underline{A}_t$. Let $\underline{X}_t$ and $\underline{x}_{it}$ be similarly defined for a covariate history, where $\mathbf{X}_t$ is the most recent set of variables that could possibly affect A_t –and are also possibly affected by past exposure, $\underline{A}_{t-1}$.

Each possible exposure history $\underline{a}$ has an associated potential outcome $Y^{\underline{a}}$. Naturally, any recipient of the program exhibits only one of these potential outcomes, the one associated to its own particular exposure history, which we assume does not depend on the exposure histories of other recipients –the stable unit treatment value assumption. That is, when some recipient is observed to have an exposure history $\underline{A}_i = \underline{a}$, then $Y_i^{\underline{a}} = Y_i$, also

expressed for all recipients and all realizations of $\underline{A}$ as $Y^{\underline{A}}|\underline{A} = Y|\underline{A}$. The other potential outcomes, the ones that were not observed because they did not actually occur, are said to be counterfactual.

With this notation, the assumption of sequential ignorability says that for any exposure history $\underline{a}$, $Y^{\underline{a}} \perp A_t|\underline{\mathbf{X}}_t, \underline{A}_{t-1} = \underline{a}_{t-1}$. It is also important to note that in order to compare the various exposure histories, IPTW estimators, as the name suggests, depend on assuming first that one has a consistent model for the probability density function (pdf) of A_t given $\underline{A}_{t-1}$ and $\underline{\mathbf{X}}_t$ (that is, the probability of A_t) and, second, at any stage t, is not the case that there is a covariate history $\underline{x}_t$ and past exposure $\underline{a}_{t-1}$ such that all recipients with such histories are certain to receive the identical exposure a_t. That is, if $f_{\underline{A}_{t-1},\underline{\mathbf{X}}_t}[\underline{a}_{t-1}, \underline{\mathbf{x}}_t] \neq 0$, then $f_{A_t|\underline{A}_{t-1},\underline{\mathbf{X}}_t}[a_t|\underline{a}_{t-1}, \underline{\mathbf{x}}_t] > 0$ for all a_t –each exposure history must have some positive probability of occurring. This last assumption is closely related to the assumption of common support.[3] Further on we discuss this problem and how to restrict the analysis to the common support.

Under this notation, the problem researchers face in estimating the causal effect of a program is that typically the potential outcomes, $Y^{\underline{a}}$, don't have the same distribution as the observed outcomes, $Y|\underline{A} = \underline{a}$; that is, $f_{Y^{\underline{a}}} \neq f_{Y|\underline{A}=\underline{a}}$, or equivalently $f_{Y^{\underline{A}}} \neq f_{Y|\underline{A}}$. This is because, in dynamic contexts, with time-varying confounders, particular levels of exposure followed recipients with particular responses to the program, which makes recipients incomparable.

However, what RHB showed is that, under the above assumptions (sequential ignorability, consistency and positivity), it's possible to generate a reweighted version of the observed outcomes that conditional on $\underline{A}$ have the same distribution as the potential outcomes using a consistent model for the pdf of A_t given $\underline{A}_{t-1}$ and $\underline{\mathbf{X}}_t$ –informally, the probability of observing the exposure history that recipient i actually took.

It is instructive to see how this works in a one stage program with a single covariate.

[3]The substance of the common support assumption is that we observe recipients with different degrees of exposure to the treatment with (roughly) the same characteristics, in this way we can base our estimations only on comparable recipients.

First note that the observed joint pdf of Y, A and X is, by definition,

$$f_{Y|A} = f_{Y^A|A} = \frac{f_{Y^A,A}}{f_A} = \frac{\int f_{Y^A,A,X}dx}{f_A} = \frac{\int f_{A|X,Y^A} * f_{X|Y^A} * f_{Y^A}dx}{f_A}$$
$$= \frac{\int f_{A|X} * f_{X|Y^A} * f_{Y^A}dx}{f_A}, \quad (2.1)$$

where the first equality follows from consistency, the second and fourth equalities follow from the definition of conditional pdf, the third from the definition of marginal pdf and the last equality follows from ignorability. Notice that if we were to use the "reweighting" function $SW(A, X) = \frac{f_A}{f_{A|X}}$ on the distribution of the observed data(Y, A, X), we would have that

$$\frac{f_A}{f_{A|X}} * \frac{\int f_{A|X} * f_{X|Y^A} * f_{Y^A}dx}{f_A} = \int f_{X|Y^A} * f_{Y^A}dx = f_{Y^A}. \quad (2.2)$$

As we can see, the reweighting function, $\frac{f_A}{f_{A|X}}$, allows us to simulate the distribution of f_{Y^A} and use the transformed data to estimate the causal effects of the program. Intuitively, as the weighting replaces $f_{A|X}$ with f_A in the observed data, the process alters the distribution of A breaking the links between the exposure and the factors that affect it.

The same principle follows with time-varying settings, as long as we assume that all confounders are observable –the idea is always to decompose the joint distribution of the observed data into an equivalent product of conditional probabilities representing different causal mechanisms, and then find the weighting function that removes from the product those factors that we find are in the way to simulate the potential outcome distribution. To demonstrate, assume that we have two covariates, X_1 and X_2, two stages with exposures, A_1 and A_2 and one outcome, Y . Assume further that the temporal order of these variables is (X_1, A_1, X_2, A_2, Y) and that every variable is affected by all its predecessors. In this case, we would have that

$$f_{Y|A_1,A_2} = f_{Y^{\underline{A}}|A_1,A_2} = \frac{f_{Y^{\underline{A}},A_1,A_2}}{f_{A_1,A_2}} = \frac{\iint f_{Y^{\underline{A}},A_1,A_2,X_1,X_2} dx_1 dx_2}{f_{A_1,A_2}}$$
$$= \frac{\iint f_{A_2|X_1,A_1,X_2,Y^{\underline{A}}} * f_{X_2|X_1,A_1,Y^{\underline{A}}} * f_{A1|X_1,Y^{\underline{A}}} * f_{X_1|Y^{\underline{A}}} * f_{Y^{\underline{A}}} dx_1 dx_2}{f_{A_2|A_1} * f_{A_1}}$$
$$= \frac{\iint f_{A_2|X_1,A_1,X_2} * f_{X_2|X_1,A_1,Y^{\underline{A}}} * f_{A1|X_1} * f_{X_1|Y^{\underline{A}}} * f_{Y^{\underline{A}}} dx_1 dx_2}{f_{A_2|A_1} * f_{A_1}}. \quad (2.3)$$

Where again we make extensive use of the definitions of conditional and marginal pdf, and sequential ignorability in the last equality. At this point the needed reweighting function suggests itself as

$$SW(\underline{A}, \underline{X}) = \frac{f_{A_2|A_1} * f_{A_1}}{f_{A_2|X_1,A_1,X_2} * f_{A_1|X_1}}. \quad (2.4)$$

Using these weights on the distribution of the observed data we would have that

$$\frac{f_{A_2|A_1} * f_{A_1}}{f_{A_2|X_1,A_1,X_2} * f_{A_1|X_1}}$$
$$* \frac{\iint f_{A_2|X_1,A_1,X_2} * f_{X_2|X_1,A_1,Y^{\underline{A}}} * f_{A_1|X_1} * f_{X_1|Y^{\underline{A}}} * f_{Y^{\underline{A}}} dx_1 dx_2}{f_{A_2|A_1} * f_{A_1}}$$
$$= \iint f_{X_2|X_1,A_1,Y^{\underline{A}}} * f_{X_1|Y^{\underline{A}}} * f_{Y^{\underline{A}}} dx_1 dx_2 = f_{Y^{\underline{A}}}. \quad (2.5)$$

This leads naturally to the stabilized weights proposed by RHB as

$$SW(\underline{A}, \underline{\mathbf{X}}) = \prod_{t-1}^{T} \frac{f_{A_t|\underline{A}_{t-1}}}{f_{A_t|\underline{A}_{t-1},\underline{\mathbf{X}}_t}}. \quad (2.6)$$

Notice that in the denominator of the weight corresponding to recipient i, SW_i, we find the probability density of the observed exposure history of that recipient conditional on the past, expressed as the product of the respective probability density at every stage, hence the name Inverse Probability of Treatment Weighting. As we have seen, the SW removes any confounding by ensuring that the distribution of exposure histories $\underline{A}$ is unrelated to the measured confounders $\underline{\mathbf{X}}$. In this way they cannot account for any remaining differences between exposure histories and, since there is no connection between the exposure

histories and the confounders in the reweighted data, one can simply run whatever model we would have used in the case of a randomized evaluation. Of course, in nonrandomized evaluations, the weights $SW(\underline{A}, \underline{\mathbf{X}})$ are unknown, and have to be estimated. The next section presents a procedure to do this.

2.3 Estimating the Stabilized Weights

To estimate the SW we need to model the exposure in each stage, conditional on the past. When exposure to the intervention comes as a single action that either occurs or not, a common approach is to estimate the probability of being exposed, $p(A_t = 1|\underline{A}_{t-1}, \underline{\mathbf{X}}_t)$, with a standard logistic regression model (logit model):

$$p(A_t = 1|\underline{A}_{t-1}, \underline{\mathbf{X}}_t; \boldsymbol{\alpha}) = (1 + exp{-}h(\underline{A}_{t-1}, \underline{\mathbf{X}}_t; \boldsymbol{\alpha}))^{-1}, \tag{2.7}$$

where h is a linear, additive function of the exposure and covariates histories, and the parameters $\boldsymbol{\alpha}$. This has been known in the literature as the *propensity score*.

In the case of continuous interventions, RHB suggest the use of Ordinary Least Squares (OLS) regression of A_t on $\underline{A}_{t-1}$ and $\underline{X}_t$ to model the distribution of the observed exposure given the past, i.e., $A_t|\underline{A}_{t-1}, \underline{X}_t \sim N(h(\underline{A}_{t-1}, \underline{\mathbf{X}}_t; \alpha), \sigma^2)$. For instance, we might have $h(\underline{A}_{t-1}, \underline{\mathbf{X}}_t; \alpha) = \alpha_0 + \alpha_1 \sum_{k=1}^{t-1} A_k + \alpha_2 \mathbf{X}_t$, which models the exposure as a function of the cumulated exposure up to the last stage and the most recent set of covariates. An estimate of the weights requires an estimate of the parameter vector $(\boldsymbol{\alpha}, \sigma)$. We can obtain these estimates,$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}^2)$, from a pooled OLS –although we could use different models for different subsets of recipients–, treating each recipient-stage as a separate observation. This would amount to use the normal distribution

$$\hat{f}_{A_t|\underline{A}_{t-1}, \underline{X}_t} = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left\{\frac{a_t - (\hat{\alpha}_0 + \hat{\alpha}_1 \sum_{k=1}^{t-1} A_k + \hat{\alpha}_2 \mathbf{X}_t)}{\hat{\sigma}}\right\}^2}, \tag{2.8}$$

as the basis for the estimation of the denominator of the weights. If we were to fit another distribution to estimate (α, σ), we would use the corresponding pdf to calculate the

weights.

For the numerator, all that is required is an additional model without conditioning on the time-varying covariates, in analogy one might specify

$$\hat{f}_{A_t|\underline{A}_{t-1}} = \frac{1}{\hat{\sigma}^*\sqrt{2\pi}} e^{-\frac{1}{2}\left\{\frac{a_t-(\hat{\alpha}_0^*+\hat{\alpha}_1^*\sum_{k=1}^{t-1}A_k)}{\hat{\sigma}^*}\right\}^2}. \tag{2.9}$$

It is important to note that there is no need to assume as Gaussian the exposure assignment model. In principle, one can estimate $f_{A_t|\underline{A}_{t-1},\underline{\mathbf{X}}_t}$ as the exposure assignment model pdf evaluated at the observed exposure and covariates histories, for the distribution and function h that better suits the program under analysis. [21] coined the term Generalized Propensity Score (GPS) for this quantity as it is analogous to the propensity score for binary coded exposures.

Once we have obtained $\hat{f}_{A_t|\underline{A}_{t-1},\underline{\mathbf{X}}_t}$ and $\hat{f}_{A_t|\underline{A}_{t-1}}$ for every recipient-stage, all we need to construct the weights is to take the product across stages and divide to obtain the estimates $\widehat{SW}(\underline{A},\underline{X})$.

In the single-shot framework a crucial diagnostic to validate the estimated propensity score consists on checking for covariate balance. We have seen in the last section that under the assumptions of sequential ignorability, consistency and positivity, the exposure is unconfounded in the weighted data, conditional on past exposure. That is, once we have broken the link between the exposure and the covariates, at any given stage of the program, the exposure might differ on a time-varying confounder, but these differences might have to do with past exposure. This implication led Blackwell [4] to suggest a balance test in a time-varying context. If after reweighting the data and conditioning on past exposure, $\underline{A}_{t-1}$, A_t is still predictive of X_t, then there is likely residual confounding of the relationship between the outcome and the exposure [4]. We can check for these associations comparing an unweighted and weighted pooled regression of each covariate X at stage t on the cumulated exposure before t, and the level of exposure at t. We would expect that the coefficient associated to this later regressor, A_t, would be statistically non-significant in the weighted data.

Also of importance in the impact evaluation literature are "common support" or "overlap region" considerations. In the single-shot framework these technical terms refer to restrict the estimation of causal effects to comparable recipients: those with similar distribution of covariates across exposures. The idea is to avoid the bias that may arise when the support of the distribution of $\mathbf{X}$ differs among groups with different degrees of exposures. In the case of fixed continuous interventions a natural way to check for common support is to split the range of the exposure in blocks and compare the distribution of the covariates among all of them –strategy proposed by Flores, Flores-Lagunes, Gonzalez and Neumann[12], FFGN from now on. It is important to keep in mind that these approaches depend crucially on the blocking on the exposure.

Note that, with fixed continuous interventions, as the number of blocks grows and ultimately reaches the number of observations in the data, comparisons end up being on individual recipients and how "far away" every one of them is from the rest in terms of the covariates. The farther away recipient i is from the rest, the less comparable it is since its particular value of X_i sets it apart from the rest of the recipients with values that no other possesses. Intuitively, it would have no comparables in the data and so is not readily useful for estimating causal effects.

Of course, there are many ways to assess the distance between one point and the rest of the data. On this matter, King and Zeng [29, 30] have proposed the use of Gower's [16, 17] metric as the basis of this assessment since this measure is design to apply to both continuous and discrete variables. The Gower distance between two points x_i and x_j, or rows in the data, is defined as the average absolute distance between the elements of the two points, divided by the range of the data:

$$G_{ij}^2 = \frac{1}{K} \sum_{k=1}^{K} \frac{|x_{ik} - x_{jk}|}{r_k}, \tag{2.10}$$

where the range is $r_k = max(X_k) - min(X_k)$ and the min and max functions return the smallest and largest elements, respectively, in the set including the kth element of the covariates X. Informally, we can see G^2 as the distance between two points expressed as

a proportion of the distance across the data –although technically speaking the distance measure is G, the square root of this quantity.

Using Gower's measure, King and Zeng [29, 30] have suggested to summarize the distances between a given point and the rest of the data as the fraction of observations in the data not farther away from the point in question than the average distance –defined as nearby observations– among all pairs of observations in the data. A similar, perhaps more natural possibility, is to use the mean distance between the recipient under scrutiny and the rest of the observations. In this fashion a mean value of G_{ij}^2 over all js equal to 0.5 means that the average distance between recipient i and all other recipients requires to travel the equivalent distance as 50% of the way across the data set. Of course, every analysis made requires a cut-off value to distinguish those recipients too far away to be considered comparable to the rest, usually based on data-conserving criteria.

Even after we have determined a cut-off value, in a time varying context –with recipient-stage observations–, it is unclear which observations to compare. Given the progressive way the weights $S\hat{W}(\underline{A}, \underline{X})$ are estimated, restrictions on the data should be imposed sequentially stage by stage, having eliminated recipients that are found to be too far away from the rest in terms of X_t before analyzing X_{t+1}, keeping in the end only those recipients whose covariates X_t are deemed "nearby" at every stage t.

It's important to note that dropping recipients outside common support changes the population of inference, restricting it, in principle, to those recipients for which we can produce good answers to our causal questions.

Once the common support has been determined, the exposure assignment model estimated and the stabilized weights validated by checking for covariate balance, we can use them to estimate the causal quantities of interest.

2.4 Mean and Quantile Dose-Response Functions for Time-varying Continuous Interventions

With time-varying interventions, estimate the causal effect of every exposure history as we would in a single-shot framework, proves a technically impossible task in general, for even with interventions that only take a handful of different degrees of intensity at every stage, the number of possible exposure histories grows geometrically with the number of stages. With a continuum of different degrees of exposures, it's nearly impossible to observe two recipients with the exact same exposure histories, let alone enough to make statistical inferences.

To estimate the effect of a time-varying intervention we need further assumptions about which exposure histories should have similar potential outcomes: a response model. These would allow for the entire history of exposure to affect the outcome in a structured, low-dimensional way. We may assume, for instance, that recipients with the same accumulated exposure should have similar potential outcomes. Whether this is a reasonable assumption or not will depend on the program under evaluation.

The substance of the evaluation, and more often than not the amount of data on hand, will determine what response model makes sense for the potential outcome. Imposing structure always comes with risks of possible model misspecification. However, if summarizing the recipients' exposure histories with a scalar function, such as the cumulative exposure $cum(\underline{A}_i) = \sum_{t=1}^{T} a_{it}$, happens to be reasonable, we could avoid making additional assumptions about the functional form of the response model using nonparametric regression methods, a strategy first suggested by FFGN in the context of fixed continuous interventions.[4] In this way, within a local linear regression framework, we could estimate a mean-response curve that maps the mean effects of program in relation to the cumulated exposure to it: a mean DRF for the cumulative exposure.

[4]FFGN proposed to estimate the mean DRF of a fixed continuous intervention using a kernel-weighted local linear regression of the outcome, Y, on the exposure, A, where each recipient's kernel weight is divided by its exposure assignment model pdf properly evaluated.

Since this pseudo-sample generated by $\widehat{SW}(\underline{A}, \underline{X})$ simulates the entire distribution of the potential outcome, it's also possible to use the same kernel smoothing methods, this time with quantile regression, to examine not just the mean, but any quantile effects of the cumulated exposure to the program and address its potential distributional effects on the outcome. In other words, we can also use the same nonparametric regression methods to estimate any quantile DRFs for the cumulative exposure.

2.5 Estimating the Effect of Seguro Popular on the Supply of Healthcare Services

Our goal in this section is to apply the procedures so far discussed to estimate the effects of the Mexican universal health insurance program, SP, on the actual capacity of the SSA to meet the new demand of healthcare services. In particular, we focused on the observed increment in the relative number (per thousand population outside the social security network) of SSA's doctors' offices, physicians and nurses in day-to-day contact with patients (providing clinical care); perhaps the most basic human and material resources involved in the provision of healthcare services. We used administrative records of the SP program containing the number of families and individuals affiliated from 2002-2013, and the federal records on infrastructure and human resources employed by the SSA over the same time period.[5]

The SP program is a clear example of a time-varying intervention. Its implementation, set out in 2002, was a process that unfolded over time and at a different pace from place to place. In fact, it took over 10 years to affiliate more than 50 million people to the program across the country, and financial resources flowed at the same pace to each state as the new

[5]The data is available at a yearly frequency disaggregated at the municipality level and can be retrieved from the Sistema Nacional de Información en Salud (National Health Information System, Sistema Nacional de Información en Salud –National Health Information System– (SINAIS)) web site: http://www.sinais.salud.gob.mx/.

health insurance covered more people.

If at every stage, the rollout of the program was affected by the density of SSA's personnel and infrastructure in its regional administrative units and, at the same time, during the years of program's expansion, the program helped build up these human and material resources, it becomes necessary to account for this dynamic selection if one is to assess the effectiveness of the program.

In our analysis we focused on the 233 Sanitary Jurisdictions (SJs) as recipients of SP coverage, defined as the number of individuals affiliated to the program relative to the size of the target population in the SJ.[6] In Mexico, SJs are the most basic administrative units of SSA in charge of the operation of healthcare services and its programs, and thus our natural unit of analysis. Note that the SP coverage attained by SJs at any given year is the result of a gradual nonstop affiliation effort that started in 2002; that is, a history of increments in the exposure to the program registered yearly in our data from 2002 to 2013.

If we were to estimate the effect of SP coverage as we would in a randomized evaluation, as if we knew that $Y^{\underline{a}} \perp \underline{A}$, we may well induce bias in the causal estimates. For it is all too likely that there are characteristics of the SJs that may have influenced the exposure history to the program that are also related to the SJs' capacity to further invest in human and material resources. It is natural to assume that SJs that exhibit a lower density of human and material resources, for one reason or another, experience difficulties to translate the federal states' financial resources into further investment in personnel and infrastructure. Had this lower density of resources been a criterion for intensifying the coverage efforts of SP, or even if in a more mechanical way the pace of the affiliation effort were determined by the SSA's personnel already on the ground, either way, particular exposure histories would have followed SJs according to their capacity to respond to the program, thus biasing our estimates.

To address these concerns, we included in our estimates of SP's effect the density of

[6] Since the SJs are comprised of several municipalities, we use the census data to approach the size of the target population as the number of individuals who lack the health insurance provided by law through the social security system to all formal workers; in other words, those outside the social security network.

resources employed by the SSA in the provision of healthcare services as time-varying covariates: doctors' offices, staffed and non-staffed hospital beds, physicians with and without day-to-day contact with patients and nurses with and without day-to-day contact with patients. It is from these 7 variables that we have determined the common support using Gower's measure as described earlier, keeping only those SJs whose average distance to the rest was less than 0.5 at every stage. Following this rule, we have kept 194 SJs, a balanced panel of 2328 SJ-year observations, discarding 17% of the data. However arbitrary this 0.5 rule may be, a notable consequence of dropping the least comparable SJs is that it makes our estimates less model dependent [30].

Tables 2.1 and 2.2 report means and standard deviations for our data before and after discarding the farthest observations.

Table 2.1: Descriptive Statistics of the Whole Sample

Variable		Mean	Std. Dev.	Min	Max	Observations
	Overall	0.08	0.10	-0.39	0.77	$N = 2796$
Exposure	Between		0.01	0.03	0.12	$n = 233$
	Within		0.10	-0.40	0.76	$T = 12$
	Overall	0.39	0.34	0.00	1.27	$N = 2796$
Cumulative exposure	Between		0.14	0.08	0.89	$n = 233$
	Within		0.31	-0.50	1.13	$T = 12$
	Overall	0.59	0.37	0.15	3.25	$N = 2796$
Doctors' offices	Between		0.36	0.16	3.02	$n = 233$
	Within		0.09	-0.07	1.98	$T = 12$
	Overall	0.65	0.86	0.00	9.70	$N = 2796$
Staffed hospital beds	Between		0.85	0.00	6.57	$n = 233$
	Within		0.18	-1.72	4.31	$T = 12$

Table continues

Variable		Mean	Std. Dev.	Min	Max	Observations
Non-staffed hospital beds	Overall	0.60	0.47	0.00	5.18	$N = 2796$
	Between		0.38	0.06	3.55	$n = 233$
	Within		0.27	-0.63	3.04	$T = 12$
Physicians in day-to-day contact with patients	Overall	1.28	1.13	0.22	11.38	$N = 2796$
	Between		1.10	0.30	10.51	$n = 233$
	Within		0.25	-1.06	3.75	$T = 12$
Physicians without day-to-day contact with patients	Overall	0.10	0.17	0.00	2.22	$N = 2796$
	Between		0.16	0.01	1.52	$n = 233$
	Within		0.05	-0.49	1.18	$T = 12$
Nurses in day-to-day contact with patients	Overall	1.65	1.60	0.11	17.03	$N = 2796$
	Between		1.57	0.28	14.29	$n = 233$
	Within		0.32	-1.20	4.38	$T = 12$

Table continues

Variable		Mean	Std. Dev.	Min	Max	Observations
Nurses without day-to-day contact with patients	Overall	0.10	0.14	0.00	1.35	$N = 2796$
	Between		0.12	0.00	1.14	$n = 233$
	Within		0.06	-0.35	1.00	$T = 12$

Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.
The table shows the standard deviation decomposed into between ($\overline{x}_i$) and within $\left(x_{it} - \overline{x}_i + \overline{\overline{x}}\right)$ components.

Table 2.2: Descriptive Statistics of the Trimmed Sample

Variable		Mean	Std. Dev.	Min	Max	Observations
	Overall	0.08	0.09	-0.20	0.67	$N = 2328$
Exposure	Between		0.01	0.05	0.12	$n = 194$
	Within		0.09	-0.21	0.67	$T = 12$
	Overall	0.38	0.33	0.00	1.21	$N = 2328$
Cumulative exposure	Between		0.12	0.15	0.77	$n = 194$
	Within		0.31	-0.39	1.12	$T = 12$
	Overall	0.50	0.22	0.15	1.81	$N = 2328$
Doctors' offices	Between		0.20	0.16	1.56	$n = 194$
	Within		0.09	-0.17	1.89	$T = 12$
	Overall	0.43	0.26	0.00	2.16	$N = 2328$
Staffed hospital beds	Between		0.24	0.00	1.39	$n = 194$
	Within		0.09	-0.09	1.19	$T = 12$

Table continues

Variable		Mean	Std. Dev.	Min	Max	Observations
Non-staffed hospital beds	Overall	0.52	0.34	0.00	2.44	$N = 2328$
	Between		0.26	0.06	1.43	$n = 194$
	Within		0.23	-0.54	1.60	$T = 12$
Physicians in day-to-day contact with patients	Overall	1.00	0.41	0.22	2.71	$N = 2328$
	Between		0.37	0.30	2.15	$n = 194$
	Within		0.19	0.14	2.04	$T = 12$
Physicians without day-to-day contact with patients	Overall	0.07	0.05	0.00	0.30	$N = 2328$
	Between		0.04	0.01	0.22	$n = 194$
	Within		0.03	-0.02	0.27	$T = 12$
Nurses in day-to-day contact with patients	Overall	1.24	0.55	0.11	3.49	$N = 2328$
	Between		0.48	0.28	2.80	$n = 194$
	Within		0.26	0.29	2.36	$T = 12$

Table continues

Variable		Mean	Std. Dev.	Min	Max	Observations
Nurses without day-to-day contact with patients	Overall	0.07	0.06	0.00	0.54	$N = 2328$
	Between		0.05	0.00	0.23	$n = 194$
	Within		0.03	-0.11	0.38	$T = 12$

Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.
The table shows the standard deviation decomposed into between $(\overline{x}_i)$ and within $\left(x_{it} - \overline{x}_i + \overline{\overline{x}}\right)$ components.

We can see from these tables that, after trimming the sample, data is more compact in every dimension, making statistical fitting more precise and, therefore, our estimates more reliable. It is worth noting that, according to the density of resources, the SJs dropped from the sample belong mostly to the upper (more resourceful) part of the distribution. Also, the within decomposition shows that the trimmed part of the sample exhibits the greater variability both upwards and downwards, but mostly upwards. Taken together, the between and within decomposition show that the SJs left out of our estimations are those better off and with higher growth rates in terms of their human and material resources. Something to keep in mind when reading the results.

Regarding our exposure assignment model to estimate RHB's stabilized weights, SW, since the observed change in the coverage of SP, A_t, is a continuous variable, we have estimated $\hat{f}_{A_t|\underline{A}_{t-1},\underline{X}_t}$ from a pooled OLS, on all 2328 SJ-year observations, using the time-varying covariates described above including all two-way interactions and past SP coverage, i.e. $cum(\underline{A}_{t-1})$, as regressors. Likewise, for $\hat{f}_{A_t|\underline{A}_{t-1}}$, we have used the same model using only $cum(\underline{A}_{t-1})$ as regressor.

After constructing $\widehat{SW}(\underline{A},\underline{X})$, we have followed Blackwell [4] looking for residual confounding comparing an unweighted and weighted pooled regression of each of the 7 covariates on the past SP coverage, $cum(\underline{A}_{t-1})$, and the increment in the coverage at t, A_t . Table 2.3 shows the results of this balance test. There we can see how coefficient associated to the change in SP coverage, A_t, turns statistically non-significant in the weighted data –the pseudo-sample generated by $\widehat{SW}(\underline{A},\underline{X})$.

Table 2.3: The Change in History-Adjusted Balance between the Weighted and Unweighted Data

		$t-statistic$	
Variable		A_t	$p > \|t\|$
Doctors' offices	Unweighted	2.59	0.010
	Weighted	0.76	0.447
Staffed hospital beds	Unweighted	0.03	0.975
	Weighted	0.41	0.679
Non-staffed hospital beds	Unweighted	2.60	0.009
	Weighted	0.07	0.944
Physicians in day-to-day contact with patients	Unweighted	3.46	0.001
	Weighted	1.20	0.228
Physicians without day-to-day contact with patients	Unweighted	-0.69	0.491
	Weighted	0.01	0.990

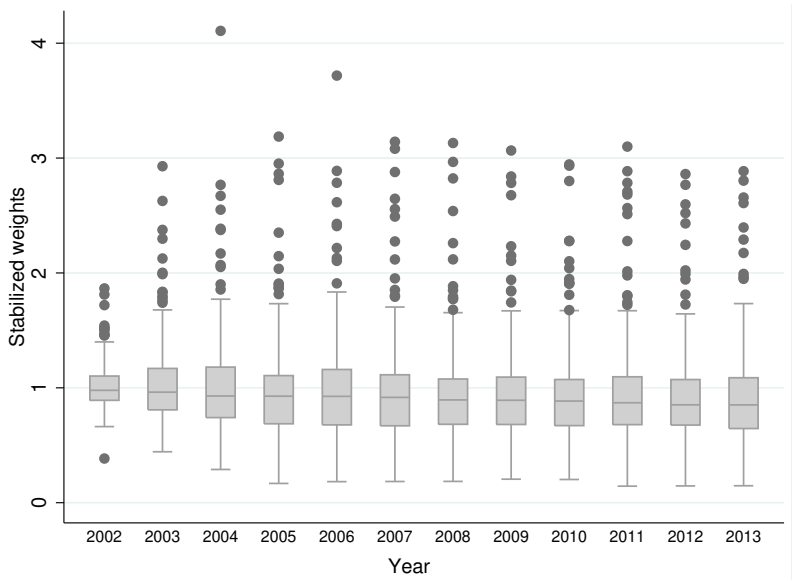
Table continues

Variable		$t-statistic$ A_t	$p > \|t\|$
Nurses in day-to-day contact	Unweighted	3.01	0.003
with patients	Weighted	0.92	0.358
Nurses without day-to-day	Unweighted	-1.71	0.087
contact with patients	Weighted	-0.23	0.820

Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

These estimates come from an unweighted and weighted pooled regression of the time-varying covariate at year t on (a) SP coverage before year t, and (b) the change in SP coverage observed in year t.

Figure 2.1: Stabilized Weights Over the Years



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

The boxes are the yearly inter-quartile ranges and the horizontal line inside of each box corresponds to the median. The whiskers represent the maximum and minimum values excluding outliers which appear as dots.

Regarding the distribution of the weights, Cole and Hernán [7] noted that the lack of common support tends to push weights away from 1. Intuitively, it would correspond to cases where the probability of certain exposure histories is close to 0 or 1 in some parts of the covariate space, which can be interpreted as violation of positivity. In Figure 2.1 we show the final distributions of the stabilized weights by year t,

$$\widehat{SW}_t(\underline{A}_t, \underline{\mathbf{X}}_t) = \prod_{k=1}^{t} \frac{\hat{f}_{A_k|\underline{A}_{k-1}}}{\hat{f}_{A_k|\underline{A}_{k-1},\underline{\mathbf{X}}_k}}, \tag{2.11}$$

where we truncate the product up until the corresponding year. There we see that the means at each year are close to 1, and that the minimum and maximum values are also reasonably close, indicating well behaved weights.

We were interested in estimating a response model for SP coverage, in other words, the cumulative exposure history $cum(\underline{A}_k) = \sum_{k=1}^{t} a_k$. Here we were assuming that,

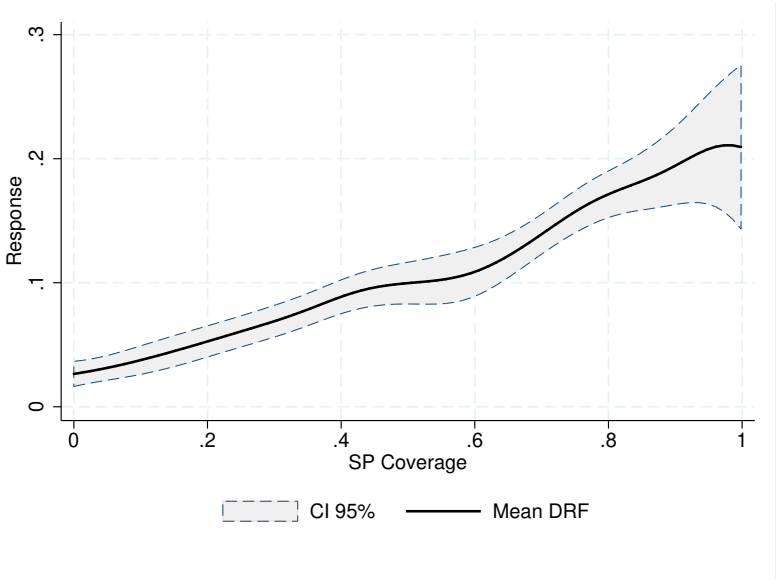
in time, the same degree of coverage attained by the SJs should have the same potential outcome, irrespective of the specific history of exposure. This seemed to be a reasonable assumption in view that the resources transferred to the states by the federal government to meet the demands of the new entitlement were directly proportional to the number of individuals affiliated to the program. In this context, attaining a particular degree of coverage early in the exposure history of a SJ would represent roughly the same financial resources as attaining it at any other point in the future. It's important to note though that it was not necessarily the case that every SJ got their corresponding share of financial resources. Regímenes Estatales para la Protección Social en Salud –State Regimes for Social Protection in Health– (REPSS) determined how and where to invest the program's financial resources. We will discuss more of this in light of our results in the conclusions.

Also, in order to draw on a larger pool of information, we assumed that the DRFs were the same for all time periods. This assumption allowed us some amount of pooling across years including all 2328 SJ-year observations in our estimates. Regarding standard errors and confidence intervals, the most straightforward way to estimate them was to bootstrap the entire estimation procedure, including the weights. Here we resampled the set of SJs and their histories, not the single SJ-year observations.

As mentioned earlier, we have followed FFGN in estimating the DRFs using a kernel-weighted local linear regression of the outcome, Y, on the cumulative exposure, $cum(\underline{A}_k)$, using the pseudo sample generated by $\widehat{SW}_t(\underline{A}_t, \underline{\mathbf{X}}_t)$. This is equivalent to multiply each SJ-year observation's kernel weight in the local linear regression by its corresponding $\widehat{SW}_t(\underline{A}_t, \underline{\mathbf{X}}_t)$. Figures 2.2 to 2.4 show the resulting mean DRFs for the observed increment in the relative number of SSA's doctors' offices, physicians and nurses in day-to-day contact with patients, with respect to 2001, the year prior the introduction of the program in its pilot phase.

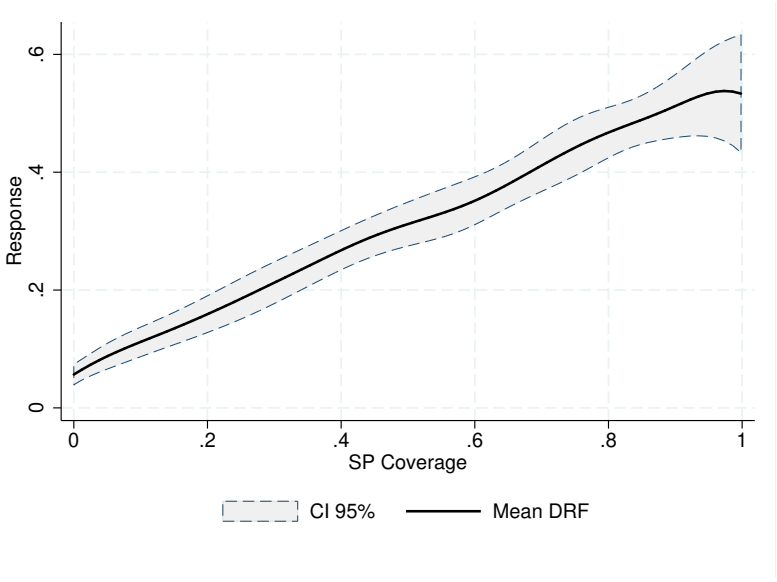
The first thing to note in these graphs is that all three response curves exhibit a positive slope, meaning that the SP program has had, on average, a positive impact on the relative numbers of these human and material resources –without a doubt a necessary condition

Figure 2.2: Increment in the Relative Number of Doctors' Offices (per thousand population outside the social security network)



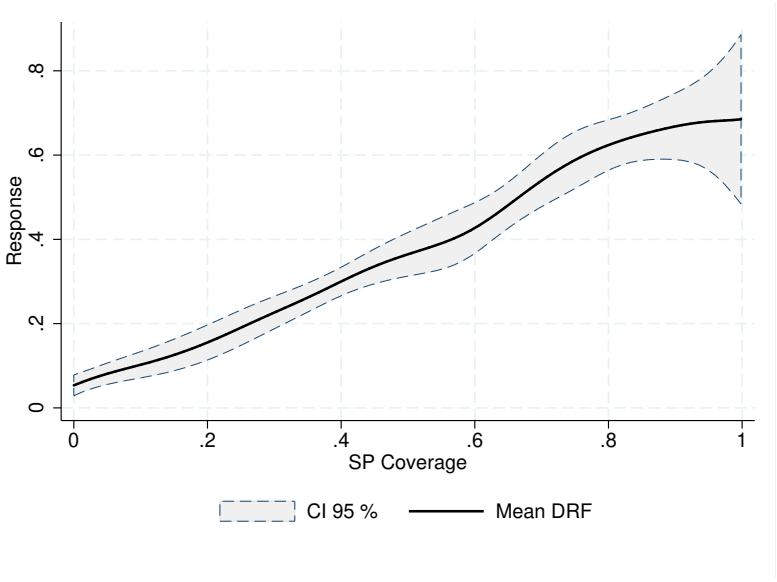
Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

Figure 2.3: Increment in the Relative Number of Physicians in day-to-day Contact with Patients (per thousand population outside the social security network)



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

Figure 2.4: Increment in the Relative Number of Nurses in day-to-day Contact with Patients (per thousand population outside the social security network)



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

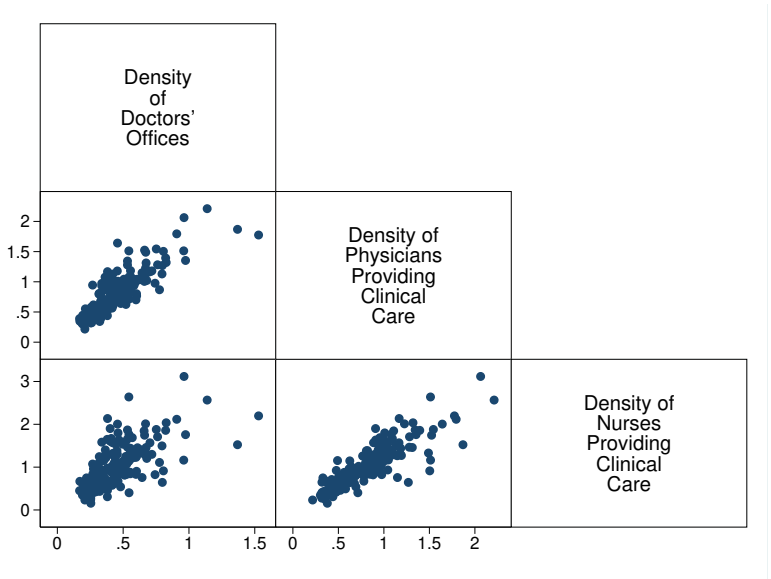
for the expansion in the provision of healthcare services. This result is in line with that of Bosch and Campos [5]. Also of importance is the difference between the end points of these mean response curves, $MeanDRF(1) - MeanDRF(0)$. Our estimates suggest that, on average, the full coverage of the SP program would represent, for the SJs in our sample, an increase in the relative number (per 1,000 people outside the social security network) of doctors' offices, physicians and nurses providing clinical care of .18, .47 and .64 respectively. Given that in 2001 the densities of these resources were .45, .81 and 1 respectively –the World Health Organization considers less than 2.3 health workers providing clinical care as a critical shortage–, we can see that the program has had an important impact on the human and material resources of the SSA.

The almost linear relationship depicted in all three graphs also suggests the absence of any economies of scale in the use of the financial resources being provided by the program. This is most striking in the case of doctors' offices where one would expect new infrastructure to be costlier. According to our estimates, the increment in the relative number of resources is, on average, directly proportional to the amount of money destined by the program per affiliate. If everything else in the program's operation had stayed the same, of course. A number to keep in mind is that in 2014 the government contribution reached 2,843.40 pesos (some 185 usd of the time) per affiliate on average, which constitutes 99% of the finances of the program [48, p.88]. It's not hard to go from here to an estimate of how much it would cost to bring the relative number of resources to a desired level if things keep running the way they have been for the last decade.

For example, in our sample of 194 SJs, in 2013 the mean relative number (per 1,000 people outside the social security network) of SSA's physicians and nurses in day-to-day contact with patients –providing clinical care– was 1.2 and 1.6 respectively. Had the government's contribution been 50% higher this year, according to our estimates, we would have seen in these SJs an average density of physicians and nurses providing clinical care of 1.4 and 1.9, respectively –"average" being the key term.

However informative these mean response curves are, they hardly provide us with a

Figure 2.5: Relative Number of Human and Material Resources 2001 (per thousand population outside the social security network)

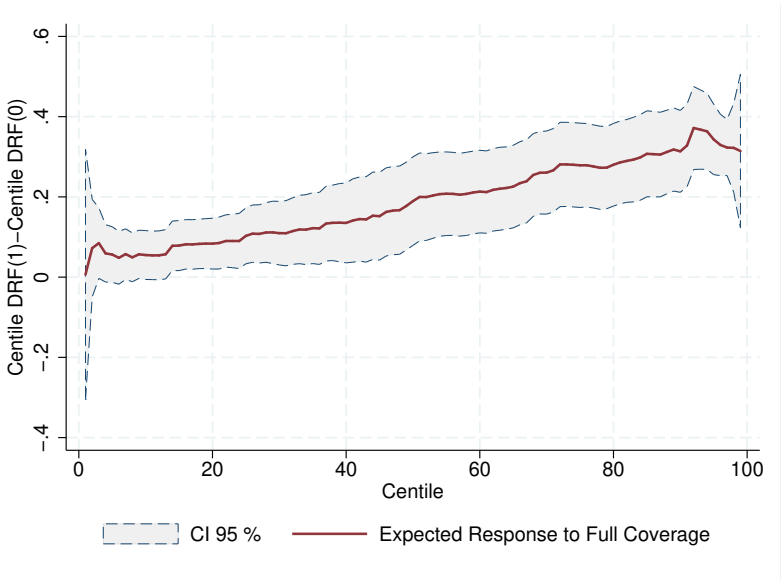


Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.
The numbers correspond to the 194 Sanitary Jurisdictions in common support.

complete description of the program's impact on other parts of the distribution of our outcome variables. This is most important for it can help us investigate the conditions under which the same coverage of the program has a differential response, possibly ascertaining subpopulations for which the program is most effective. In general, profiling the recipients that are making the most of a program is something of great importance, for it can help improve its design and steer it in the desired direction. We can see the relevance of this exercise for the SP program looking at the dispersion of our data the year prior the introduction of the program in Figure 2.5.

Figure 2.5 shows the great differences one could find in 2001 looking at the density of resources in the SJs. Heterogeneous responses to the program among this variability is to be expected, even more so if what we observe in 2001 is the result of historical

Figure 2.6: Expected Increment in the Relative Number of Doctors' Offices due to Full Coverage of SP (per thousand population outside the social security network)



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

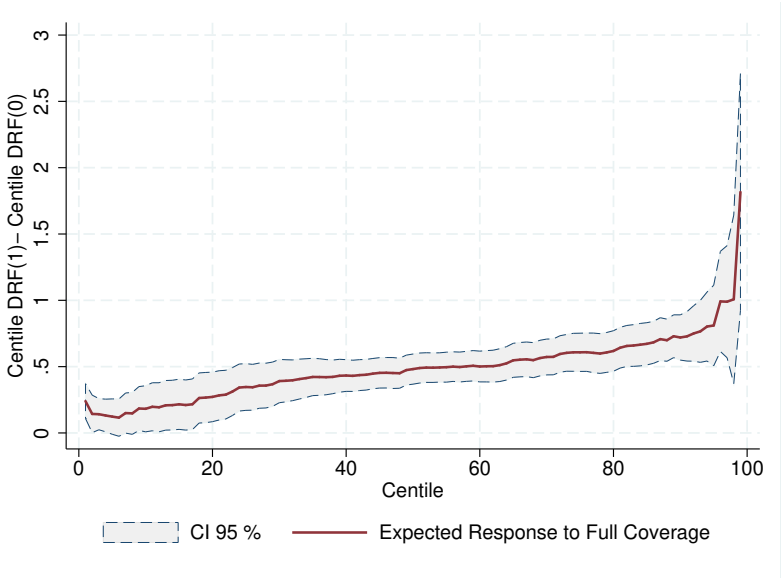
obstacles that the SJs face to translate the state's financial resources into further investment in personnel and infrastructure.

As previously stated, we can also estimate quantile versions of the same response curves using the same kernel-weighted regression framework, only this time with linear quantile regressions of the outcome on the cumulative exposure. This will give us a clear understanding of how the program is affecting the whole distribution of our outcome variables, taking us one step further into characterizing the determinants of possible heterogeneous responses to the program.

Figures 2.6 to 2.8 show the estimated response of full coverage of SP on every centile of the distributions of the outcome variables; that is, the vertical difference between the end points of every centile i response curve, $Centile_i DRF(1) - Centile_i DRF(0)$.

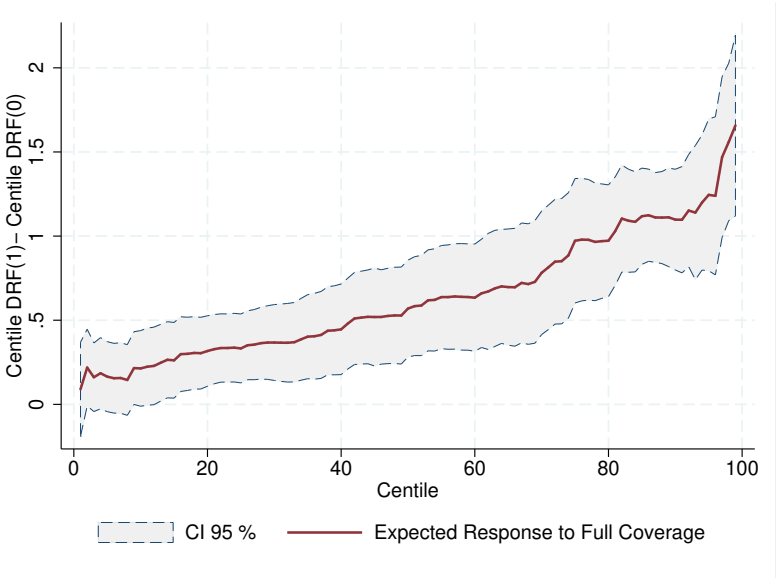
All three figures show a positive tendency for the expected response to full coverage

Figure 2.7: Expected Increment in the Relative Number of Physicians in day-to-day Contact with Patients due to Full Coverage of SP (per thousand population outside the social security network)



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

Figure 2.8: Expected Increment in the Relative Number of Nurses in day-to-day Contact with Patients due to Full Coverage of SP (per thousand population outside the social security network)



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

of the program. The fact that these figures do not exhibit horizontal lines suggests that going from zero to full coverage of SP has different responses along the distribution of the outcome variables, in this case, showing greater effects on the upper part of the distributions. In other words, the better the SJ is doing increasing the relative number of its human and material resources, the greater its response to the program. Note how this is not a tautology, it might as well could have been just the opposite. What this means is that the program gives greater boosts to those SJs increasing the relative number of its resources the most –again almost in a direct proportion.

There are at least two reasons why this might be happening. The first is that this could be a simple reflection of the SJs' investment priorities. It may well be that those SJs exhibiting greater increases in the relative number of their resources find within their investment priorities such expansion, and the program's financial resources help them to do exactly just that. If this were the case, it is only natural to observe that those SJs whose investment priorities lie elsewhere different from expanding the density of their human and material resources do not use the program's money in this fashion.

A second reason for the positive slope in figures 2.6 to 2.8 is that it might be reflecting historical obstacles the SJs face to transform financial means into human and material resources, obstacles that the program by itself cannot overcome. In this second scenario, we would be looking not so much at the result of different investment priorities in themselves, but to technical difficulties in transforming the programs budget into medical staff and infrastructure. It is well known, for example, that it is costlier to invest in some areas due to limited geographical access. This alone would naturally lead to lesser responses to the program in SJs where the need for resources is dearer. We may even be looking at the result of a compromise between equality and efficiency, which always implies an ethical dilemma. We'll discuss more of this in the final section when we highlight the public policy implications.

In light of Figures 2.6 to 2.8, given that SP seems to be having distributional effects on the outcome variables, it becomes relevant to determine where SJs fall in the distribution

of our outcome variables, since this determines which SJs are making the most of the program in terms of the availability of resources needed in the provision of healthcare services. Finding out whether the main effects of a social program are felt among the least or most vulnerable part of the population leads to very different appreciations of its impact. Of course, these are not the only two possible scenarios but the extremes of a continuum of possibilities.

A thorough investigation of what is actually happening will require the development of new procedures to estimate counterfactual distributions under different possible scenarios, a pending assignment for time-varying continuous interventions. However, as an exploratory analysis we can take a look at how the 2001-2013 increments in human and material resources were distributed across SJs according to their initial level. In particular, Figures 2.9 to 2.11 show to what extent it can be said that the increment in these resources has been pro-poor in terms of resource in question.[7]

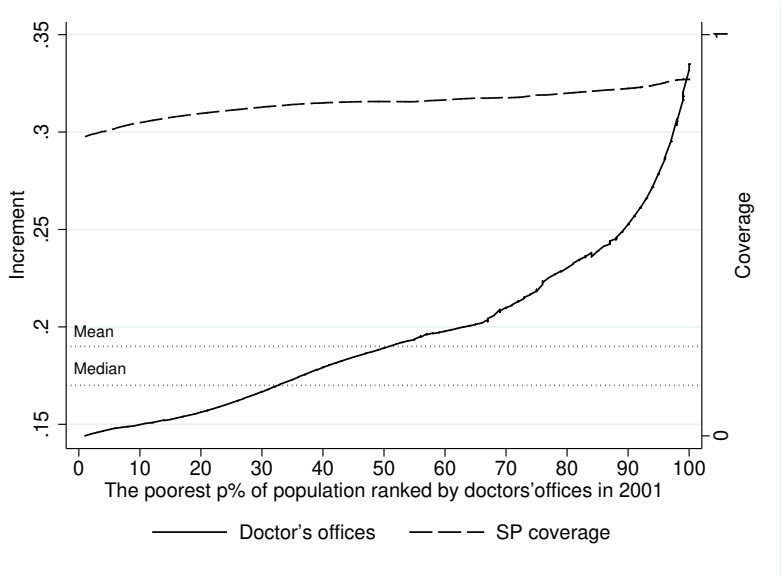
As can be seen from this last set of figures, all of our estimates suggest that the greatest increments observed in the relative number of human and material resources from 2001 to 2013 corresponded to those who were better off in 2001. We have also plotted the mean SP coverage corresponding to every centile to show that this pattern is hardly explained by differences in the rollout of the program. Absolute indices of inequality capture the evolution of this result over the 12-year period. Figure 2.12 shows the change in the absolute GINI indices[8] for the material and human resources of the SSA. Given that the absolute GINI index is not unit invariant we have normalized all indices so that 2002=100.

Together, the increment incidence curves and inequality indices suggest that the SJs making the least of the SP program are precisely those in greater need, for these are also

[7] All estimates were weighted by the population outside the social security network.

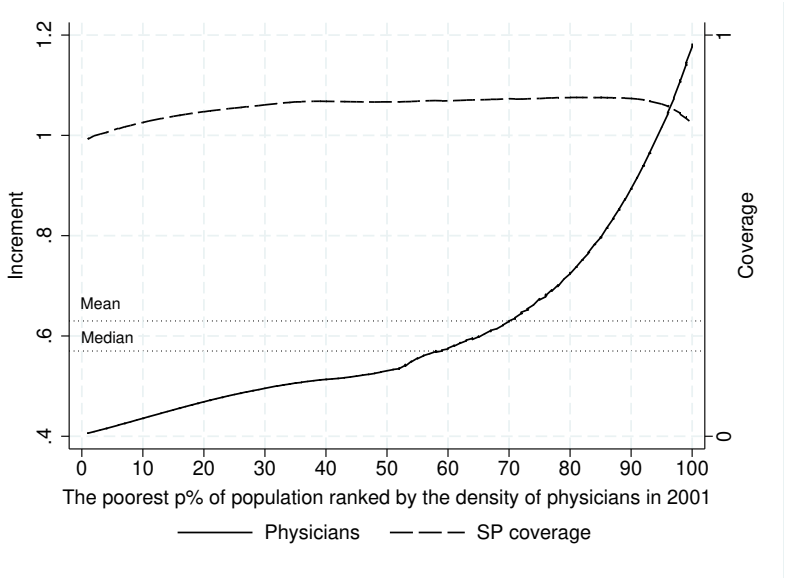
[8] In the absolute version of the GINI index inequality is unaffected by an equal addition to all incomes (translation invariant). In the standard (relative) version of the GINI index inequality is unchanged when all incomes are increased in the same proportion (scale invariant). The absolute criterion was advocated by Kolm [34, 35]. Several Kolm indices were also estimated showing the same tendency. The difference between these the relative and absolute measures is that the absolute criterion always considers improvement in terms of absolute difference.

Figure 2.9: Increment Incidence Curve for the Relative Number of Doctor's Offices, 2001-2013 (per thousand population outside the social security network)



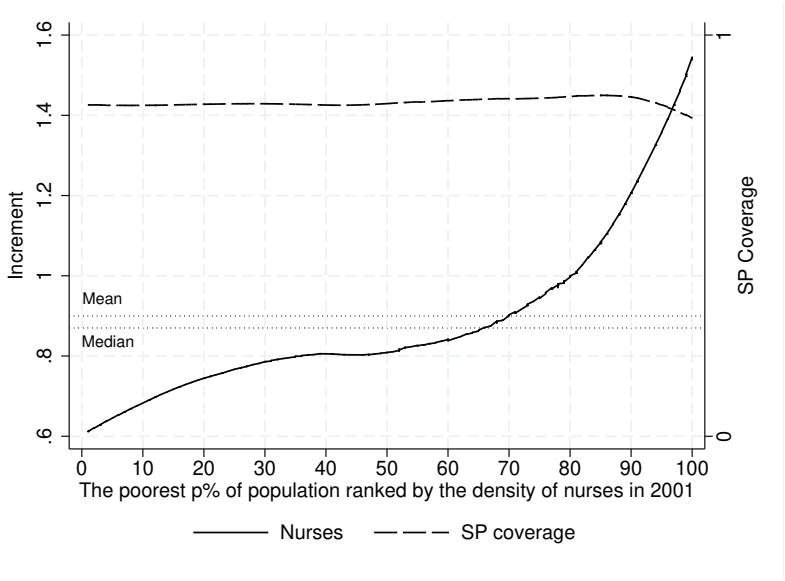
Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

Figure 2.10: Increment Incidence Curve for the Relative Number of Physicians in day-to-day Contact with Patients, 2001-2013 (per thousand population outside the social security network)



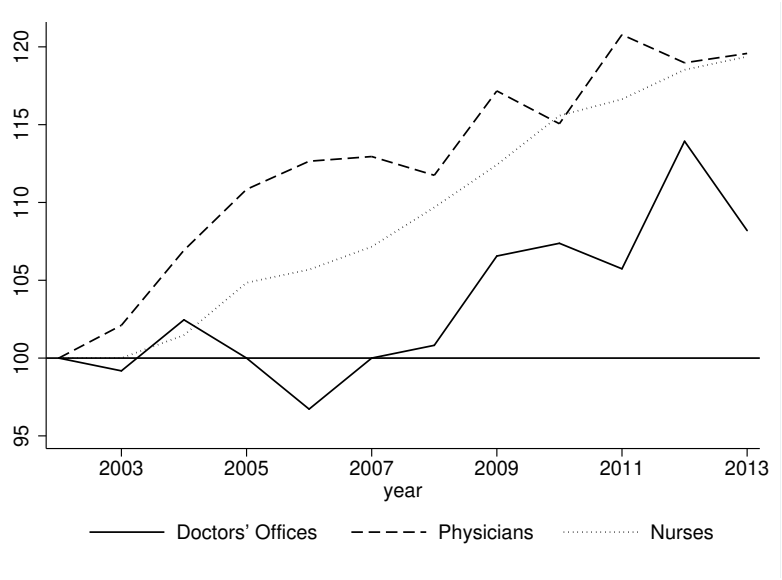
Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

Figure 2.11: Increment Incidence Curve for the Relative Number of Nurses in day-to-day Contact with Patients, 2001-2013 (per thousand population outside the social security network)



Source: Personal elaboration based on data from the National Health Information System (SINAIS) and the administrative records of Seguro Popular.

Figure 2.12: Inequality in SSA Human and Material Resources, 2001-2013: Absolute GINI (Indices: 2002=100)



Source: authors' elaboration on data drawn from the National Health Information System, SINAIS. Estimates from grouped data weighted by the population outside the social security network at the level of Sanitary Jurisdiction.

the ones showing the least increment in basic human and material resources at roughly the same coverage of the program. This should be a matter of major concern for any program directed at vulnerable population groups.

2.6 Conclusions

Impact evaluation, like any other scientific inquiry, requires some background assumptions. However, assumptions must be made carefully, for it is well known that assumptions way too off the mark can generate wildly inaccurate conclusions. We argue that this is exactly what would happen if one were to apply single-shot causal inference methods to estimate the impact of social programs (treatments) whose interventions do not take place all at once as single homogeneous action, but instead unfold over time with different intensity across recipients according to the milieu themselves help create as they unfold.[9]

To control for the time-depending confounding inherent to the dynamic nature of these kinds of programs, in this article we have brought together two strands of the literature on causal inference in observational studies: one focusing on fixed non-binary treatments and the other on binary dynamic treatments. The procedure we present here elaborates on [43] showing how to estimate mean and quantile dose-response functions of continuous dynamic treatments much in the same fashion as Flores et al. [12] did for continuous fixed treatments; that is, applying local regression methods to appropriately weighted samples.

We have shown the potential of this approach analyzing the effects of the Seguro Popular program in Mexico on key variables associated with the provision of healthcare services by the SSA. Unlike previous evaluations, we have found compelling quantitative evidence that the program has proven most helpful in less vulnerable territories, leaving behind those in greater need.

Regarding the palpable differences in the results the SP has shown across the country,

[9] If, for example, we had ignored the dynamic nature of our empirical application, using only pre-program covariates in an attempt to control for confounders, we would have overestimated the mean impact of the program by approximately 45%.

the authorities behind the program have blamed the lack of accountability in the financially decentralized system within which it operates, as well as the autonomy with which every federal state determines how to best invest their corresponding financial resources [9].

Beyond the problems inherent to the lack of accountability, it is a well-known fact that rural and urban areas present quite different challenges for the provision of health services. In this sense, our findings might be due in part to efficiency concerns in the investment decisions at the federal entity level, for it's all too likely that we don't see in rural areas the economies of scale obtained in urban areas [39]. However, concerns with equity are also an important element of programs aimed at improving social conditions, and the pursuit of efficiency must never eclipse equity considerations in this context. Without a doubt, whatever trade-off between equity and efficiency there might be in the allocation of health care resources, given its importance, deserves an open discussion.

Lastly, it remains a pending task to explore the possibility to simulate counterfactual distributions that could shed more light on the distributional effects of dynamic continuous interventions. This would give us a better sense of what distributional effects we are to expect from social programs such as Seguro Popular.

Chapter 3

Robust Estimation of Inverse Probability Weights for Time-Varying Continuous Interventions

In this Chapter we present a continuous extension for longitudinal analysis settings of the recently proposed Covariate Balancing Propensity Score (CBPS) methodology. While extensions of the CBPS methodology to both marginal structural models and general treatment regimens have been proposed, these extensions have been kept separately. We propose to bring together this previous work using the generalized method of moments to estimate inverse probability weights such that after weighting the association between time-varying covariates and the treatment is minimized.

3.1 Introduction

Propensity Score (PS) methods are considered by many to be the least preferred quasi-experimental impact evaluation method (say, vis-a-vis regression discontinuity) because their strong identifying assumptions –more importantly that treatment assignment is ignorable conditional on observed confounders–, particularly in settings where those eligible to participate may chose otherwise. However, propensity score methods can be applied to a wide range of different research settings, and it is this versatility that has resulted in the popularity of PS estimates as a tool for making causal inference.

As shown by Robins [41], PS methods might even be used to evaluate programs that evolve over long periods of time, reacting to changes in the environment that themselves help to bring about. That class of statistical models came to be known as Marginal Structural Models (MSMs) [43, 42, 4]. Despite its theoretical appeal, MSMs as a tool for making causal inference from longitudinal data have yet to extend to general treatment regimens. Surprisingly, while several extensions of propensity score methods to general treatment regimens have been proposed [27, 43, 26, 21, 12] and keep gaining popularity among applied researchers [52], these methodologies haven't been coupled with the MSM framework.

Perhaps behind this gap between methodological and applied research is the practical difficulty of estimating the inverse probability weights –required for MSMs to generate a pseudo sample from which to estimate the casual quantities of interest–, whose correct model is generally unknown and its misspecification can actually increase bias, even if the selection on observables assumption holds. In contrast with cross-sectional PS methods, in MSMs the effect of model misspecification is exacerbated, for it propagates across time since the algorithm provided by Robins, Hernán and Brumback [43] to estimate the weights in these models involves the product of propensity scores estimated for each time period.

Since, in theory at least, inverse probability weighting renders the intervention inde-

pendent of time-varying covariates, usual practice involves iteratively checking if weighting on the estimated probability meets this expected property –revising the model until said correlation is minimized. The downside of this algorithm is that it can lead to ad hoc specifications of the probability models which may not agree with the preferred theoretical specification.

As a way to get around this practical difficulty surrounding MSMs, Imai and Ratkovic [25] came up with the idea of formally dovetailing this correlation minimizing property of the weights into the model estimation method itself, eliminating the need to manually and iteratively check the treatment assignment model. They accomplished this by making use of the Generalized Method of Moments (GMM), adding to the model's score vector –that is, the gradient of the log-likelihood– the correlation minimizing property as an extra moment condition. They have called this new approach the Covariate Balancing Propensity Score (CBPS) [24] and it can easily be implemented through publicly available open-source software [14]. Simulation and empirical studies suggest that the CBPS improves the empirical performance of MSMs by making the treatment assignment model more robust to misspecification. However, so far the CBPS extension to MSMs has been confined to binary time-dependent treatments.

Researchers facing continuous time-dependent treatments may be tempted to dichotomize the exposure at every time period in order to make use of the existing CBPS estimator for MSMs and its advantages, if at the cost of compromising data analysis through the loss of valuable information. Here we dispense with this temptation proposing a further extension to the CBPS methodology. Our proposal builds on the recent generalization of the CBPS methodology to general treatment regimens [13] taking it to the longitudinal setting, thus bridging the gap between two extensions of the original CBPS estimator.

In Section 3.2 we briefly review MSMs and their assumptions. Section 3.3 we describe the proposed extension to the CBPS methodology. In Section 3.4 we present our simulation study. In addition, we present an empirical application in Section 3.5 to show the performance of the proposed methodology with real data. Finally, we offer concluding

comments in Section 3.6.

3.2 Marginal Structural Models

After introducing some notation, in this Section we briefly review the MSM framework [43, 42, 4].

Assume a program that evolves over long periods of time intervening with different intensity across recipients at every stage of their implementation, reacting to changes in the environment which the program itself contributed to bring about; that is to say, reacting to its own previous results in a feedback loop. In this context, variables that can both affect future exposure to the program and be affected by past exposure may confound the impact of the program in a way that no cross-sectional impact evaluation method can sort out [4].

Notation

Denote recipient i's observed k-dimensional vector of time-varying covariates at stage t of the program $\boldsymbol{X}_{it}$, with $i = 1, \ldots, N$ and $t = 1, \ldots, T$. At every stage t of the program, recipients are observed receiving the exposure level a_{it} to the benefits of the program; that is, one possible realization of the exposure variable A_{it}, with $a_{i0} = 0$. Collecting all the observed exposures to the program for a given recipient, from its launch up to stage t, gives us the history of exposure $\underline{a}_{it} = (a_{i1}, \ldots, a_{it})$. Independent, identically distributed realizations of $(Y_{it}, \underline{A}_{it}, \underline{\boldsymbol{X}}_{it})$ are observed for each recipient at every stage of the program, where Y_{it} is some outcome variable of interest. Let $\underline{\boldsymbol{X}}_{it}$ and $\underline{\boldsymbol{x}}_{it}$ be similarly defined for a covariate history, where $\boldsymbol{X}_{it}$ is the most recent set of variables that could affect A_{it} –and are also affected by past exposure history, $\underline{A}_{it-1}$.

Following Robins' adaptation of the potential outcome framework [49, 46] to longitudinal settings [40], we use $Y_{it}^{\underline{a}_t}$ to represent the potential outcome for recipient i measured at time t associated to the exposure history $\underline{a}$ up to t, which we assume does not depend on

the exposure histories of other recipients. This is often referred to in the literature as the stable unit treatment value assumption. Naturally, any recipient of the program exhibits only one of these potential outcomes at any given stage t; that is, the one associated to its own particular exposure history up to that point in time $Y_{it} = Y_{it}^{\underline{a}_{it}}$, the consistency assumption. All other, unobserved, potential outcomes are said to be counterfactual.

As with other PS methods, MSMs are based on the assumption of no unmeasured confounding, in this case at each time period. This assumption, also referred to as sequential ignorability [4] or conditional interchangeability [42], states that the exposure level at stage t is statistically independent (ignorable) of the potential outcomes, conditional on the covariate and exposure histories up to that point. Using previous notation, sequential ignorability says that for any exposure history up to t, a_t, $Y_{it}^{\underline{a}_t} \perp A_{it} \mid \underline{A}_{it-1} = \underline{a}_{t-1}, \underline{\boldsymbol{X}}_{it} = \underline{\boldsymbol{x}}_t$. Additionally, also typical in PS methods, MSMs also assume that at any stage t, there is no covariate history $\underline{\mathbf{x}}_t$ and past exposure $\underline{a}_{t-1}$ such that all recipients with such histories are certain to receive the identical exposure a_t. That is, each exposure history must have some positive probability of occurring. This last assumption can be formally written as, if $f\left(\underline{A}_{it-1}, \underline{\mathbf{X}}_{it}\right) \neq 0$ then $f\left(\underline{A}_{it} \mid \underline{A}_{it-1}, \underline{\mathbf{X}}_{it}\right) > 0$, where f stands for the probability density function (pdf).

Under the above assumptions (sequential ignorability, consistency and positivity), Robins [41] showed that the conditional probability density of exposure history can be used to consistently estimate the impact of such dynamic interventions. In this generalization of Inverse Probability of Treatment Weights (IPTW), the probability of observing a particular exposure history is expressed as the product of the respective probability density at every stage, $\prod_{t-1}^{T} f\left(A_t \mid \underline{A}_{t-1}, \underline{\mathbf{X}}_t\right)$. However, using this probability directly as weights when fitting the outcome model leads to highly variable estimates and researchers usually follow the suggestion given in the literature [7] and use the stabilized weights of the form

$$SW_t\left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right) = \prod_{t'=0}^{t} \frac{f\left(A_{t'}\right)}{f\left(A_{t'} \mid \underline{A}_{t'-1}, \underline{\boldsymbol{X}}_{t'}\right)} \tag{3.1}$$

where, for convenience, we define $SW_0 = 1$. In nonrandomized evaluations, irrespective

of the nature of the intervention whose impact one might be trying to estimate –whether it's best coded as binary, multivalued or continuous–, these weights are unknown, and have to be estimated. Usually, a parametric model is used to estimate the probability density of exposure given the past at each time period. When the intervention comes as a single action that either occurs or not at every stage of the program (i.e. coded as a binary variable), a common approach is to estimate this probability density at every stage with a standard logistic regression model (logit model). In the case of continuous interventions, Gaussian models (with constant variance) are often assumed, i.e., $A_t|\underline{A}_{t-1}, \underline{\boldsymbol{X}}_t \sim N\left(h\left(\underline{A}_{t-1}, \underline{\boldsymbol{X}}_t; \boldsymbol{\gamma}\right), \sigma^2\right)$, which can be estimated by the GMM. If there is reason to believe that h is a linear, additive function of exposure and covariates histories, and parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)$, researchers may even regress the current level of exposure on all past exposure and covariates,[1] e.g.,

$$f_\theta\left(A_t \mid \underline{A}_{t-1}, \underline{\boldsymbol{X}}_t\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[A_t - \left(\underline{A}_{t-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_t^\top \boldsymbol{\beta}\right)\right]^2\right\}. \tag{3.2}$$

The numerator of the stabilized weights is typically estimated using the empirical distribution of the exposure. If these later assumptions hold, and further assuming the marginal distribution of A_t to be Gaussian with mean zero (possibly after the pertinent transformation), the stabilizing weight is given by[2]

$$\begin{aligned} &SW_{it}\left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right) \\ &\qquad = \prod_{t'=0}^{t} \exp\left[\frac{1}{2\sigma^2}\left\{-2A_{it'}\left(\underline{A}_{it'-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta}\right) + \left(\underline{A}_{it'-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta}\right)^2\right\}\right]. \end{aligned} \tag{3.3}$$

Once these stabilized weights are estimated, all the researchers have to do to is use them in running whatever model of the outcome, as function of treatment history alone, they would have used in the case of a randomized evaluation.[3]

[1] The Least square estimator can also be viewed as a special case of GMM estimator.

[2] Equations are derived in detail in the appendix.

[3] For $t = 1$, these weights are exactly the same as those proposed by [13] for continuous treatments in cross sectional settings.

3.3 Correlation Breaking IPTW

In this Section we describe our proposal for a procedure to estimate IPTW for MSMs with continuous treatments robust to misspecification.

As with other PS methods, Robins' stabilized weights break the link between the exposure history and the time-varying covariates, guaranteeing the exposure history is unconfounded. That is, at any given stage of the program, we expect the histories of both exposure and the time-varying covariates to be uncorrelated in the weighted data [4, 23].

Formally, under our time-varying setting, we can see this correlation-breaking property for all t as

$$\mathrm{E}\left[SW_{it}\left(\underline{A}_{it}, \underline{\boldsymbol{X}}_{it}\right)\underline{A}_{it}\underline{\boldsymbol{X}}_{it}\right] = \mathrm{E}\left(\underline{A}_{\mathrm{it}}\right)\mathrm{E}\left(\underline{\boldsymbol{X}}_{\mathrm{it}}\right). \tag{3.4}$$

Note that if we center both the exposure variable and the covariates, equation 3.4 equals zero. We propose to use this weighted cross moment as an extra moment condition in a GMM framework, which for t=1 is the same Fong, Hazalett and Imai[13] used in the cross-sectional setting.

Including equation 3.4 among the score conditions –the gradient of the log-likelihood with respect to $\boldsymbol{\theta}$– as moments, allows to account for both properties we would expect the stabilized weights meet in a single estimator: predict the exposure among recipients and break the link between the exposure and the time-varying covariates. There are many ways to specify the moment conditions for the GMM estimator depending on the assumptions one is willing to make in estimating the probability models and how to include the correlation-breaking condition, each with different computational burden.

Following a common practice of assuming $A_t \mid \underline{A}_{t-1}, \underline{\boldsymbol{X}}_t$ independent and identically distributed across time periods, and making use of the fact that it also follows from the correlation-breaking property that current exposure is uncorrelated with the past, we have the following moment conditions,

$$g_\theta\left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right) = \begin{pmatrix} \frac{1}{\sigma^2}\left(A_{it} - \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta}\right)\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \\ -\frac{1}{2\sigma^2}\left\{1 - \frac{1}{\sigma^2}\left[A_{it} - \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta}\right)\right]^2\right\} \\ \prod_{t'=0}^{t} \exp\left[\frac{1}{2\sigma^2}\left\{-2A_{it'}\left(\underline{A}_{it'-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta}\right) + \left(\underline{A}_{it'-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta}\right)^2\right\}\right] A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \end{pmatrix}.$$

Note that the score conditions correspond to the first order conditions of the log-likelihood

$$l\left(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \underline{A}_t, \underline{\boldsymbol{X}}_t\right) = \frac{1}{2} \sum_{i=1}^{N} \left\{ \log\left(2\pi\sigma^2\right) + \frac{1}{\sigma^2} \left[A_{it} - \left(\underline{A}_{it'-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^{\top} \boldsymbol{\beta} \right) \right]^2 \right\}. \quad (3.5)$$

We have conveniently rearranged the histories of exposure and covariates in the correlation-breaking condition for symmetry's sake, the proof of which follows the same general outline as before.

The GMM estimator of $\boldsymbol{\theta}$ is obtained minimizing

$$\hat{\boldsymbol{\theta}}_{CBSW} = \operatorname{argmin}_{\boldsymbol{\theta}} \overline{g}_{\boldsymbol{\theta}} \left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right)^{\top} \sum\nolimits_{\boldsymbol{\theta}} \left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right)^{-1} \overline{g}_{\boldsymbol{\theta}} \left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right), \quad (3.6)$$

where $\overline{g}_{\boldsymbol{\theta}}$ is the mean over i of $g_{\boldsymbol{\theta}}$, and $\sum_{\boldsymbol{\theta}} \left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right)$ is the mean over i of the 3×3 covariance matrix of entries

$$\sum \boldsymbol{\theta}[1,1] = \frac{1}{\sigma^2}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top} \tag{3.7}$$

$$\sum \boldsymbol{\theta}[1,2] = \sum \boldsymbol{\theta}[1,2] = 0 \tag{3.8}$$

$$\sum \boldsymbol{\theta}[1,3] = \sum \boldsymbol{\theta}[3,1] = \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top} \tag{3.9}$$

$$\sum \boldsymbol{\theta}[2,2] = \frac{1}{2\sigma^4} \tag{3.10}$$

$$\sum \boldsymbol{\theta}[2,3] = \sum \boldsymbol{\theta}[3,2] = \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)}\left\{\frac{-\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)}{\sigma^2}\right\}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \tag{3.11}$$

$$\begin{aligned}\sum \boldsymbol{\theta}[3,3] = & \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2}\left\{\sigma^2 + \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)^2\right\} \\ & \exp\left\{\frac{\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)^2}{\sigma^2}\right\}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}.\end{aligned} \tag{3.12}$$

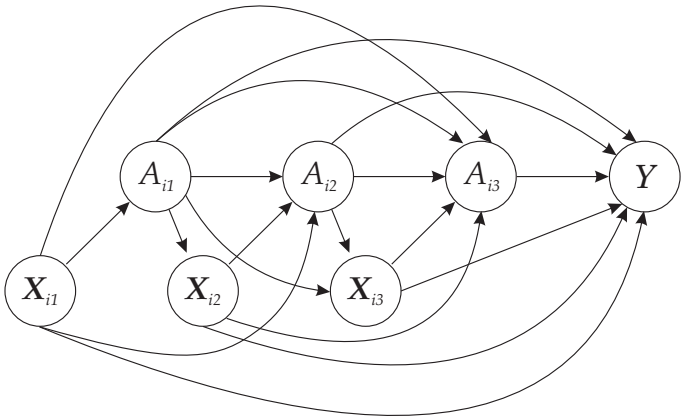
The upper left 2×2 submatrix of $\sum_{\boldsymbol{\theta}}\left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right)$ is well-known since corresponds to the covariance of the score conditions of the normal distribution. The rest of the elements are derived in detail in the appendix.

3.4 Simulation Studies

In this section we show the results of simulation studies in order to assess the empirical performance of these correlation-breaking stabilized weights.

We consider the case of three time periods, i.e., $t = 1, 2, 3$, and use three different sample sizes $n = 300$, 900 and 1,500, corresponding to 100, 300 and 500 exposure-recipients across our time frame. In our setup, summarized in Figure 3.1, the treatment-generating process is a function of a vector of exogenous time-varying covariates, $\underline{\boldsymbol{X}}_{it}$, and the previous history of exposure, $\underline{A}_{it-1}$. The outcome variable, Y, is a function of both entire histories of covariates and exposure.

Figure 3.1: Data Generating Process in Simulation Study



We use this Data-Generating Process (DGP) to examine the performance of the proposed methodology, vis-à-vis the usual Ordinary Least Squares (OLS) approach, when faced with both misspecification of the treatment-assignment model –incorrectly specifying the lag structure– and measurement error.

We use a similar simulation setup as Imai and Ratkovic [25]. The simulation consists of four covariates; for time t, we use the covariates

$$\boldsymbol{X}_{it} = (Z_{it1} \cdot U_{it}, Z_{it2} \cdot U_{it}, |Z_{it3} \cdot U_{it}|, |Z_{it4} \cdot U_{it}|) \tag{3.13}$$

where each Z_{itk} is an $i.i.d.$ draw from the standard normal distribution, and U_{it} is constructed as

$$U_{it} = \prod_{t'=1}^{t} (1 + .4A_{it'-1}) \tag{3.14}$$

for $t = 2, 3$ and $U_{it} = 1$ for $t = 1$. The continuous treatment A_{it} is generated by

$$A_{it} = \sum_{t'=1}^{t} \left\{ \left(\frac{1}{2}\right)^{t'-1} (\boldsymbol{\alpha} \boldsymbol{X}_{it'} + A_{it'-1}) \right\} + v_i \tag{3.15}$$

where $\alpha = (.8, -.2, .4, -.6)^\top$ and v_i is a standard normal disturbance. All parameters were selected in an almost random manner looking only to guarantee a minimum dispersion. The continuous outcome Y is generated by a linear combination of the whole exposure and confounders histories such that

$$Y_{it} = 275 + 10 \sum_{t'=1}^{t} A_{it'} + \boldsymbol{\delta} \sum_{t'=1}^{t} \boldsymbol{X}_{it'} + u_i \tag{3.16}$$

where $\delta = (-23.5, 12.5, 7.3, -17.4)^\top$ and u_i is another normal disturbance with mean zero and standard deviation five. Note that we have generated the outcome variable for every t, making it a function of both entire histories of covariates and exposure. We haven't included these elements in the DGP depicted in Figure 3.1 in order not to over-complicate the diagram.

In the misspecified scenario, we use the following non-linear transformation of the covariates

$$X_{it} = \left(X_{it1}^3, \exp\left(X_{it2}\right), \log\left(X_{it3}\right), \frac{1}{X_{it4}} \right) \tag{3.17}$$

as measurement error, as well as incorrectly specify the lag structure of the treatment-assignment model using only the data from the immediately previous period such that

$$A_{it} = \boldsymbol{\alpha} \boldsymbol{X}_{it} + A_{it-1} + v_i, \tag{3.18}$$

but maintain the correct outcome model.

We report results of one thousand simulated datasets for each scenario. First, with every dataset, we fit a Pooled OLS model (GLM) as the treatment assignment model, using

correct and incorrect model specifications as discussed above. Then we fit the same models using the proposed correlation-breaking IPW methodology (CBIPW) in two ways: first using both score and correlation-breaking conditions (Over-identified CBIPTW) as described in section 3.2, and with the correlation-breaking condition alone (Just-identified CBIPW), ignoring the score conditions. This gives us three different estimates according to the importance placed on both expected properties of the weights. The Generalized Linear Model (GLM) estimates obviate the correlation-breaking property, ignoring the subvector g_3, while the just-identified CBIPW pays no attention to predict exposure assignment, ignoring the subvector $(g_1, g_2)^\top$. Over-identified CBIPW stands in the middle attending to both criteria. Finally, we regress in a pooled model the outcome variable on the cumulated exposure, $\sum_{t'=1}^{t} A_{it'}$ using the stabilized MSM weights. The coefficient that results from this last regression is compared with the numerical estimate obtained from a thousand simulations using large datasets ($n = 30{,}000$) and the true exposure assignment probabilities.

Table 3.1 presents the results for our three sample sizes. The columns show the bias and Root Mean Squared Error (RMSE) from the pooled weighted linear regression of the outcome on the cumulated exposure using the stabilized weights.

The results are in line with those of previous Covariate Balance Propensity Scores for MSMs [25]. When the exposure assignment model is correct specified, all methods have small bias and small RMSE. In contrast, when the model is misspecified, the bias and RMSE are large and even grow in sample size. However, the RMSE of the CBIPW estimates grow at a much smaller rate, thereby outperforming OLS.

In short, Table 3.1 tells a simple story for time-varying continuous interventions: if researchers are lucky enough to know the true DGP, they are probably better off with OLS. If, on the other hand, researchers have reasonable doubts with respect to the origin of their data, they are probably better off with CBIPW.

Table 3.1: Performance of Probability Weights Estimation Methods

	$n = 300$		$n = 900$		$n = 1500$	
Estimator	Bias	RMSE	Bias	RMSE	Bias	RMSE
Correct Specification						
True	1.19	3.79	0.70	3.38	0.47	3.09
GLM	1.19	3.61	0.70	3.23	0.46	2.96
CBIPW Over-identified	1.44	3.94	1.11	3.01	0.99	2.90
CBIPW Just-identified	1.47	3.93	1.07	3.06	0.98	2.85
Misspecification (measurement and lag structure)						
GLM	4.87	8.66	3.38	15.04	3.58	24.41
CBIPW Over-identified	3.89	6.42	3.65	7.27	3.77	8.66
CBIPW Just-identified	4.16	6.04	3.84	6.72	3.99	7.63

3.5 Empirical Application

As an empirical application of the proposed method, we estimate the impact of the Mexican universal health insurance program, Seguro Popular (SP), on key variables associated with the provision of healthcare services by the Secretaría de Salud –Mexican Ministry of Health– (SSA). Here we analyze the same data of Chapter 2; that is, the federal records on infrastructure and human resources employed by the SSA,[4] and the coverage records of the program at a yearly frequency from 2001, the year prior the introduction of the program, to 2013.

Specifically, data is aggregated to the Sanitary Jurisdiction (SJ) level, the most basic administrative units of SSA in charge of the operation of healthcare services and its programs, and time-varying covariates include doctors' offices, staffed and non-staffed hospital beds, physicians and nurses, both with and without day-to-day contact with patients. Following the original analysis, 39 out of 233 SJs are dropped so as to make the common support

[4] Available from the National Health Information System's web site: http://www.sinais.salud.gob.mx/.

assumption more credible, and the estimates less model dependent.

The original estimations fit a single probability model, regressing the yearly increment of SP coverage on the most recent set of covariates (per thousand population outside the social security network) along with all their two-way interactions, and past (cumulated) SP coverage. In contrast, here we drop all the interactions from the model.

Looking at the degree of covariate balance achieved by the CBIPW (Fig. 3.2, Appendix Table A.4), we see that it is improved in all cases, as medians are closer to zero, and that is best achieved using the just-identified approach, both on location and dispersion as expected. It is worth noting that, in this case, fitting only the probability model (GLM) does not seems to break the correlation between the exposure to treatment and three covariates: doctors' offices, physicians and nurses in day-to-day contact with patients.[5]

Using weighted linear regressions to estimate the impact of (cumulated) SP coverage on the change in the (relative) numbers of doctors' offices, physicians and nurses (providing clinical care), we can see (Fig. 3.3 , Appendix Table A.9) that in all probability the effect of SP is positive in all three outcome variables, and grows bigger the less correlated are the covariates with the treatment in the pseudo sample. While these results are somewhat similar to those presented in Chapter 2, every point estimate here is marginally bigger (by 10% for doctors' offices, and roughly 3% for physicians and nurses).

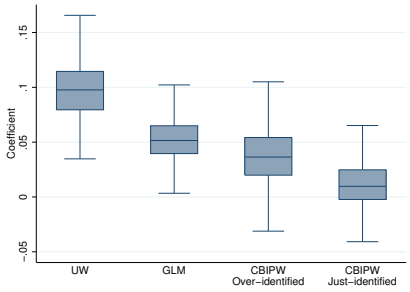
3.6 Conclusions

We have shown that the idea behind the Covariate Balancing Propensity Score (CBPS) [24] is flexible enough to cover also longitudinal settings with general treatment regimens. Here we provide a robust estimation method for MSMs with continuous treatments whose main advantage is that it directly includes the correlation-breaking property of Robins' [41] stabilized weights in the estimation procedure. This new tool avoids the manual

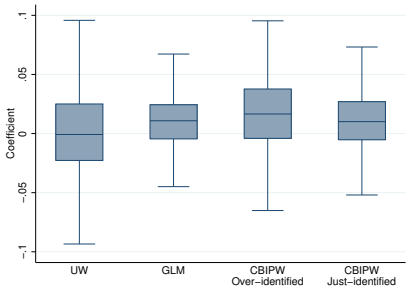
[5]Even though in the original analysis an OLS model is fitted, we do not observe this imbalance in the covariates because a different model was fitted with an extra 21 variables.

Figure 3.2: Covariate-Balance for Time-varying SP coverage on Sanitary Jurisdictions

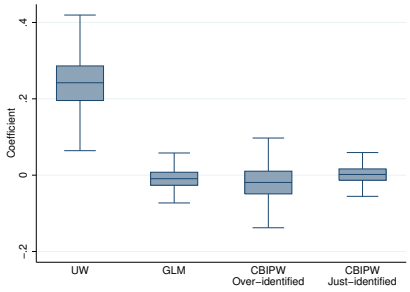
(a) Doctors' offices



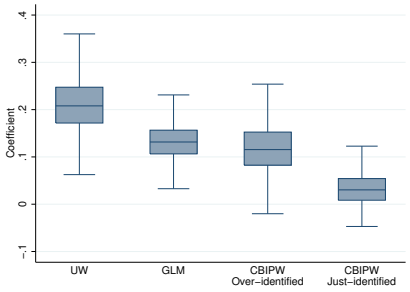
(b) Staffed hospital beds



(c) Non-staffed hospital beds

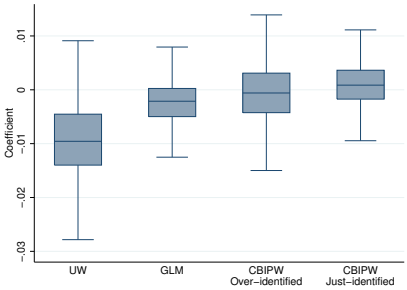


(d) Physicians*

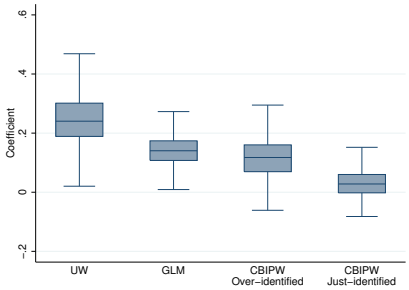


(*Figure continues*)

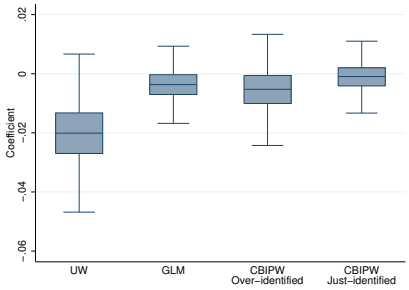
(e) Physicians**



(f) Nurses*



(g) Nurses**



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records.

The bottom and top of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Coefficients result from pooled linear regressions of each covariate on the yearly increment of SP coverage, weighted accordingly to the pseudo-sample in question.

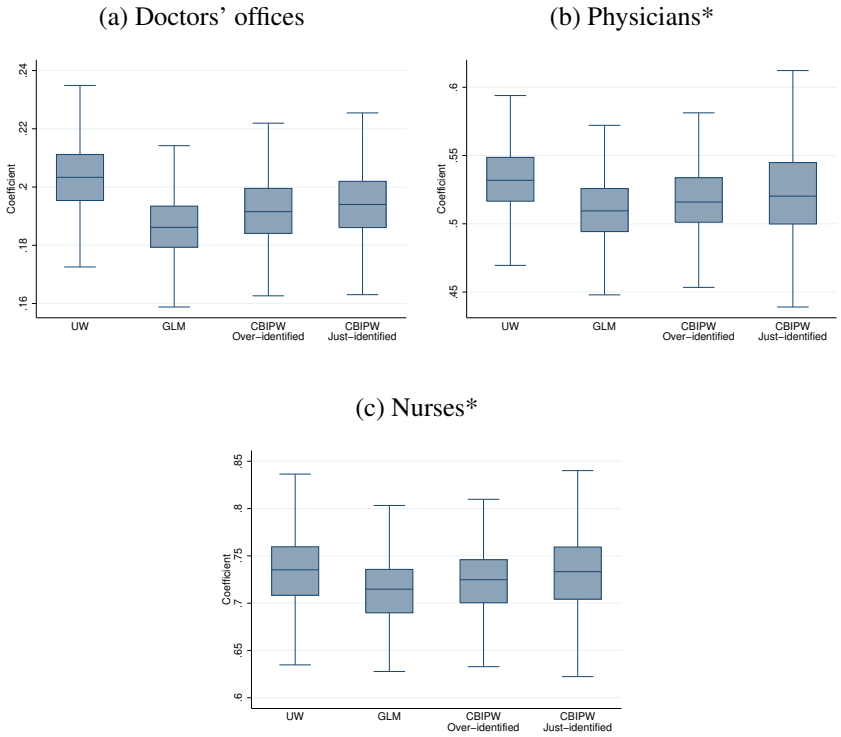
Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

* In day-to-day contact with patients

SP coverage refers to affiliates as proportion of the population without traditional health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

UW corresponds to the unweighted 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data; GLM to the sample weighted using OLS and an empty model as stabilizer; CBIPW-OVER to the sample weighted using over-identified correlation-breaking inverse-probability weights, which gives equal importance to both correctly predict treatment-assignation and balancing covariates and CBPS-EXACT to the sample weighted using Just-identified CBIPW, which privileges covariate balance over the probabilistic model.

Figure 3.3: Average Treatment Effects of SP coverage on Sanitary Jurisdictions



Source: Personal elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records.

The bottom and top of each box are the first and third quartiles, and the band inside is the second quartile (the median). Whiskers represent the lowest datum still within 1.5 Inter Quartile Range of the lower quartile, and the highest datum still within 1.5 Inter Quartile Range of the upper quartile. Outside values are not shown.

Coefficients result from pooled linear regressions (2328 obs.) of each outcome variable on SP coverage, weighted accordingly to the pseudo-sample in question.

Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

* In day-to-day contact with patients

SP coverage refers to affiliates as proportion of the population without traditional health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

UW corresponds to the unweighted 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data; GLM to the sample weighted using OLS and an empty model as stabilizer; CBIPW-OVER to the sample weighted using over-identified correlation-breaking inverse-probability weights, which gives equal importance to both correctly predict treatment-assignation and balancing covariates and CBPS-EXACT to the sample weighted using Just-identified CBIPW, which privileges covariate balance over the probabilistic model.

process of checking the correlation amongst the current exposure to treatment and the covariate histories in the weighted data and then respecifying. We are confident that this new extension to the CBPS helps clear the way for those interested in causal analysis in these complex settings.[6]

It can be argued that minimizing the correlation between covariates can be done efficiently with already available methods, such as the non-parametric extension of the CBPS to general treatment regimens [14] or Genetic Matching [10]. However, unlike these other methods, the methodology we propose here provides the researcher also with a robust probability model from which it is possible to start building bridges between MSMs and decomposition methods literatures, something that remains a pending task for future research.

[6] Stata code is available upon request from the authors.

Final Comments

The research here presented is both methodological as is empirical: analyzing the effects of Seguro Popular (SP) on the medical infrastructure and human resources in Mexico, we have pushed forwardPropensity Score (PS) methods. Regarding our findings, everything suggests that the Mexican universal health insurance program is leaving behind the most vulnerable geographic areas in the country. Not because the population there lack insurance, but because the effect of SP has not made itself felt in all regions. For a program specifically intended for the poor, this is of concern and calls for better designs or complementary efforts. The procedures and estimators here advanced helped reaching this conclusions, and hopefully will continue helping others in their own causal investigations for better, more effective, public policies in benefit of those in greater need.

Resumen (Spanish)

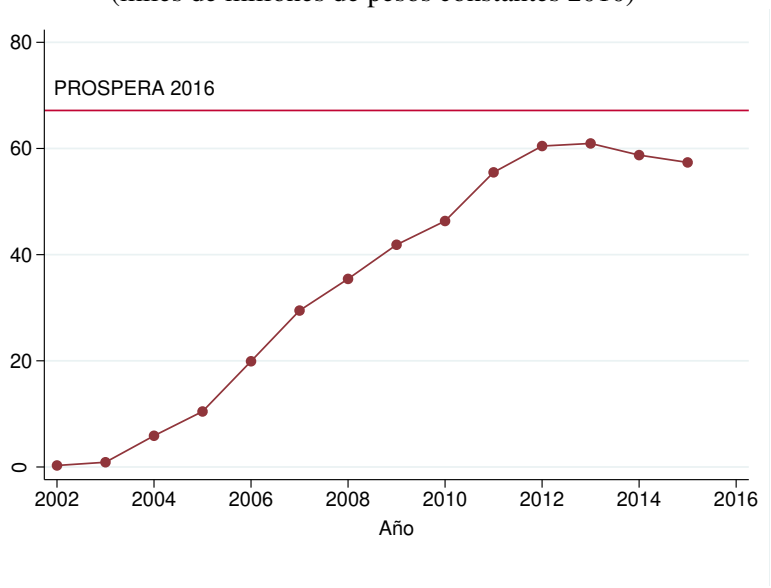
Nadie que requiera servicios de salud debería quedarse sin recibirlos por falta de recursos. Este imperativo ético ha sido recogido como derecho constitucional en casi todos los países de América Latina incluyendo a México. El Sistema de Protección Social en Salud –System of Social Protection in Health– (SPSS), mejor conocido como el Seguro Popular (SP), es el esfuerzo más significativo que México ha hecho por garantizar el derecho a la protección de la salud de aquellos que no cuentan con la seguridad social que acompaña al empleo formal, prácticamente la mitad del país. A finales de 2015, el SP registraba más de 57 millones de personas afiliadas, y un presupuesto programado para 2016 que rebasa los 75 mil millones de pesos, equivalente al 92 % del presupuesto del Programa de Inclusión Social/Oportunidades (Fig. E.1).

Simultáneamente, la Secretaría de Salud –Mexican Ministry of Health– (SSA) ha ampliado los recursos físicos y humanos involucrados en la prestación de los servicios amparados por el SP. De 2001 a 2015, el número de Establecimientos de Salud ha aumentado 31 %, la mayoría (1 de cada 3) Centros de Salud Rurales de 1 núcleo básico de servicio (típicamente conformado por un médico y una o dos enfermeras), Casas de Salud (sede de brigadas móviles que visitan las comunidades), Unidades de Especialidades Médicas (UNEMES) y Unidades Móviles (Fig. E.2).

En ese mismo periodo, los consultorios médicos han aumentado 61 %. Una cifra que, si bien es correlato del incremento de Centros de Salud Rurales y UNEMES, ha sido im-

Figura E.1: Gasto ejercido por el Seguro Popular

(miles de millones de pesos constantes 2010)



Fuente: elaboración propia con datos de la Secretaría de Hacienda y Crédito Público
http://finanzaspublicas.hacienda.gob.mx/es/Finanzas_Publicas/Cuenta_Publica

pulsada de manera importante por los Hospitales (1 de cada 3 de los nuevos consultorios lo registran Hospitales Generales, Especializados e Integrales; véase Fig. E.3).

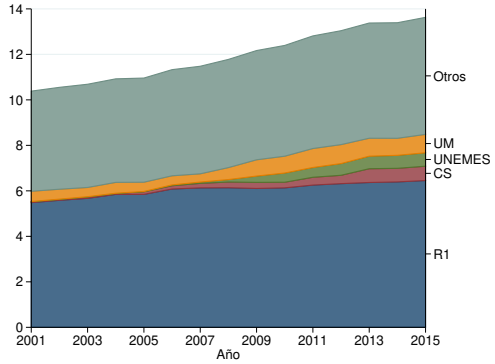
La alta prioridad política de los hospitales también se expresa en su participación del incremento de médicos y enfermeras en labores clínicas (Fig. E.3). En 2015 el número de médicos en contacto con pacientes creció 48 % con respecto a 2001, más de 33 mil nuevos médicos, de los cuales 7 de cada 10 están adscritos a hospitales. Lo mismo ocurre con las enfermeras, si bien estas se han más que duplicado desde 2001.

Basta una mirada a la evolución de los egresos hospitalarios de estas instituciones para hacerse una idea de la nueva demanda que enfrenta la SSA (Fig. E.4). Este número se ha más que duplicado desde 2001, y en 2015, 7 de cada 10 de estos egresos corresponde a personas afiliadas al SP.

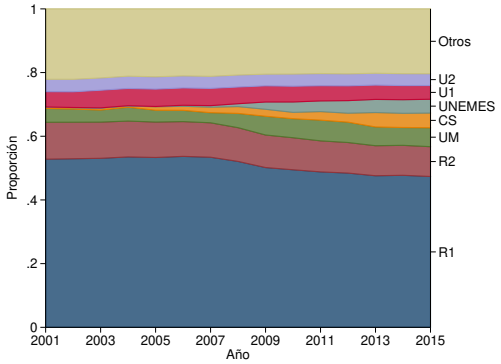
Sin embargo, no todo el país se ha beneficiado igual de la expansión de los recursos de la SSA. El incremento en la densidad –por cada mil habitantes sin seguridad social

Figura E.2: Establecimientos de la Secretaría de Salud por tipología

(a) Miles de establecimientos



(b) Proporción



Fuente: elaboración propia con datos de la Dirección General de Información en Salud

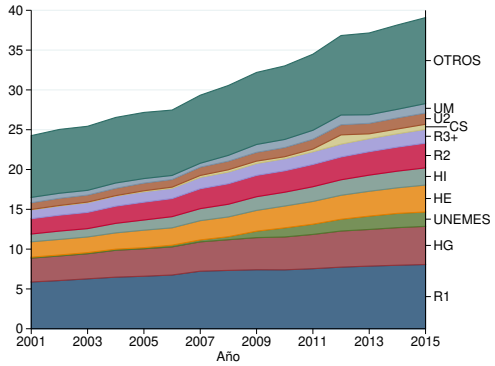
Las tipologías se ordenan, en el panel *a* de acuerdo a su participación en el incremento total de establecimientos de 2001 a 2015, y en el panel *b* de acuerdo a su participación en el total de 2015, con las de mayor participación en la base.

Donde R1 hace referencia a los Centros de Salud Rurales de 1 núcleo básico de servicio, CS a Casas de Salud, UNEMES a Unidades de Especialidades Médicas, UM a Unidades Móviles, U1 y U2 a Centros de Salud Urbanos de 1 y 2 núcleos básicos de servicio respectivamente y en Otros se agregan el resto de tipologías incluyendo Almacenes, Antirrábicos, Brigadas Móviles, Centros de Salud con Hospitalización, Centros Avanzados de Atención Primaria a la Salud, Centros de Salud con Servicios Ampliados, Clínicas de Especialidades, Consultorios Delegacionales, Hospitales Especializados, Hospitales Generales, Hospitales Integrales, Hospitales Psiquiátricos, Oficinas Administrativas, Establecimientos de Apoyo, Centros de Salud Rurales de 2 y 3 o más Núcleos Básicos, Unidades del Ministerio Público y Centros de Salud Urbanos de 3, 4, 5, 6, 7, 8, 9, 10, 11, y 12 o más Núcleos Básicos.

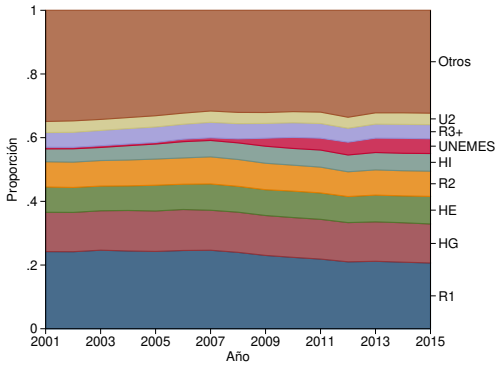
Figura E.3: Recursos de la Secretaría de Salud por tipología de establecimiento

Consultorios

(a) Miles



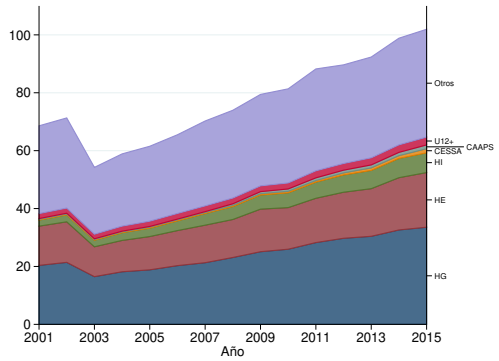
(b) Proporción



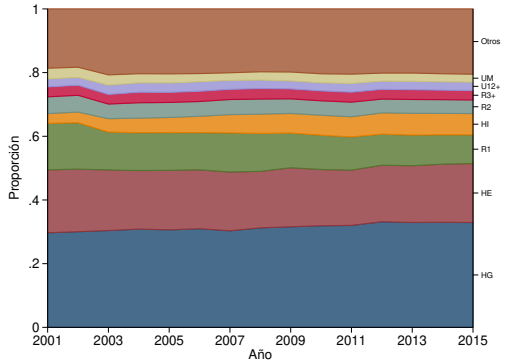
(continúa)

Médicos en contacto con pacientes

(c) Miles



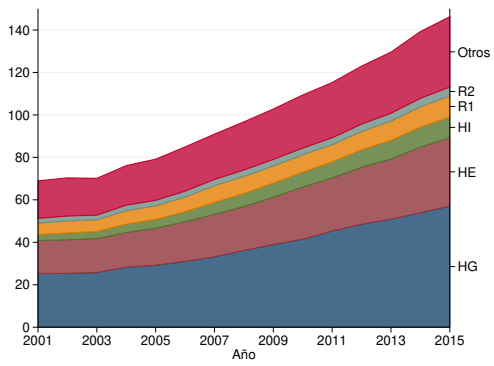
(d) Proporción



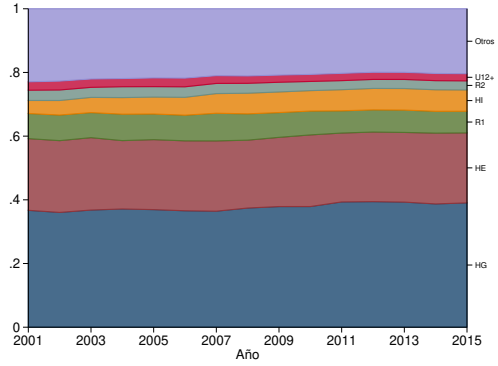
(continúa)

Enfermeras en contacto con pacientes

(e) Miles



(f) Proporción

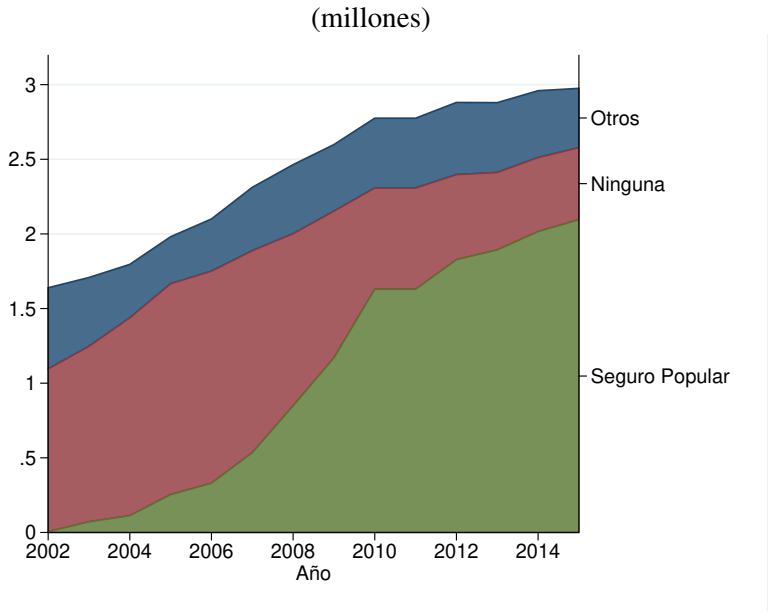


Fuente: elaboración propia con datos de la Dirección General de Información en Salud

Las tipologías se ordenan en las gráficas, a la izquierda, de acuerdo a su participación en el incremento total de 2001 a 2015 y, a la derecha, de acuerdo a su participación en el total de 2015, con las de mayor participación en la base.

R1, R2 y R3+ hace referencia a Centros de Salud Rurales de 1, 2 y 3 o más núcleos básicos de servicio respectivamente, HG a Hospitales Generales, HE y HI a Hospitales Especializados e Integrales respectivamente, UNEMES a Unidades de Especialidades Médicas, CS a Casas de Salud, U2 a Centros de Salud Urbanos de 2 núcleos básicos de servicio, UM a Unidades Móviles, CESSA a Centros de Salud con Servicios Ampliados, CAAPS a Centros Avanzados de Atención Primaria a la Salud, U12+ a Centros de Salud Urbanos de 12 o más núcleos básicos, y Otros agrupa al resto de tipologías que incluye Almacenes, Antirrábicos, Brigadas Móviles, Centros de Salud con Hospitalización, Clínicas de Especialidades, Consultorios Delegacionales, Hospitales Psiquiátricos, Oficinas Administrativas, Establecimientos de Apoyo, Unidades del Ministerio Público, Centros de Salud Urbanos de 1, 3, 4, 5, 6, 7, 8, 9, 10 y 11 núcleos básicos.

Figura E.4: Egresos hospitalarios de la Secretaría de Salud según derechohabiencia al Seguro Popular



Fuente: elaboración propia con datos de la Dirección General de Información en Salud

Se cuentan únicamente aquellos egresos registrados en establecimientos de salud a cargo de la Secretaría de Salud

tradicional (asociada al trabajo): Intituto Mexicano del Seguro Social (IMSS), Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE), Petróleos Mexicanos (PEMEX), Defensa o Marina– de consultorios, médicos y enfermeras en labores clínicas varía notablemente entre las jurisdicciones sanitarias del país (Fig. E.5).

Si bien la evolución de la densidad de recursos físicos y humanos en cada jurisdicción sanitaria cuenta su propia historia, es posible distinguir estadísticamente tres tipos de trayectorias de crecimiento para cada uno de estos recursos; donde, en cada caso, los mayores incrementos se observan en las jurisdicciones sanitarias con mayor densidad de recursos. Esto es que, las zonas del país con menos médicos, enfermeras y consultorios por habitante son las que menos incrementos han visto desde la puesta en marcha del SP.

Desde su concepción, el SP se ha presentado como una política de combate a la pobreza, focalizada en la población vulnerable, lo que obliga a preguntarnos por el impacto que el programa ha tenido sobre la falta y desigual distribución de los recursos destinados a la protección de la salud.

Dado el tipo de intervención que opera el SP, dar respuesta a la pregunta por el efecto del SP sobre la distribución de los recursos físicos y humanos de la SSA ha requerido unir dos ramas de investigación sobre evaluación de impacto: la inferencia con variables causales continuas por un lado y con tratamientos dinámicos por el otro.

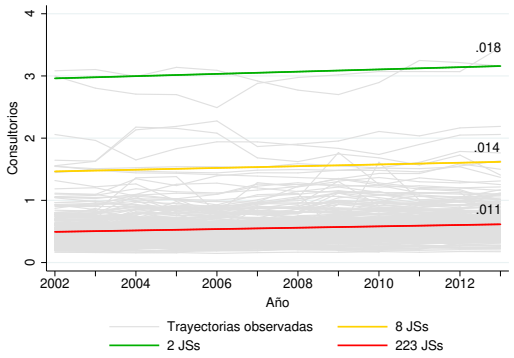
El periodo de tiempo tan largo que ha tomado el despliegue completo del programa, y su probable retroalimentación, ha demandado el desarrollo de un procedimiento semiparamétrico para analizar intervenciones continuas variables en el tiempo, así como de un estimador que facilite la evaluación de este tipo de intervenciones. Esto último a través de un programa estadístico, disponible gratuitamente para la comunidad académica, escrito en el lenguaje de la paquetería estadística Stata® (véase Apéndice B). La metodología estadística desarrollada permite perfilar grupos de beneficiarios según su respuesta a intervenciones complejas como la del SP, un adelanto cuya potencia desborda el análisis llevado a cabo aquí y que habrá de probarse útil en contextos similares.

La flexibilidad de la metodología adoptada sugiere explorar varias vetas de investi-

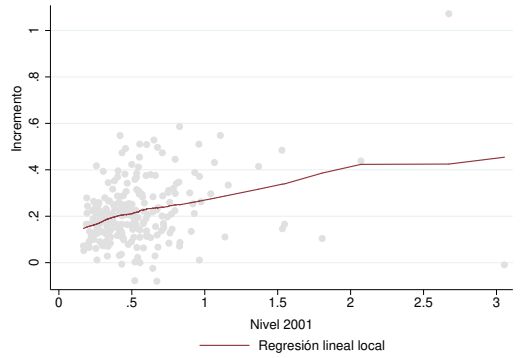
Figura E.5: Densidad de recursos de la Secretaría de Salud por jurisdicción sanitaria

Consultorios

(a) Clases de trayectorias



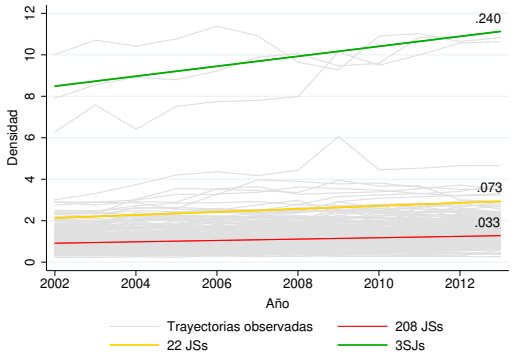
(b) Diferencia 2013-2001 según niveles de 2001



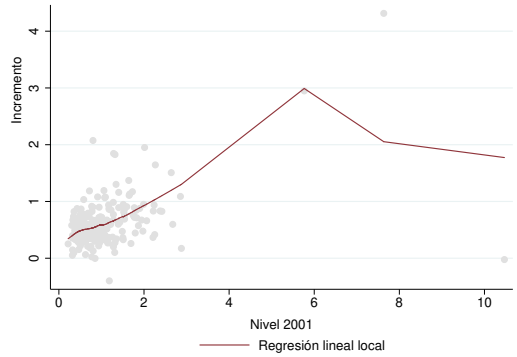
(continúa)

Médicos en contacto con pacientes

(c) Clases de trayectorias



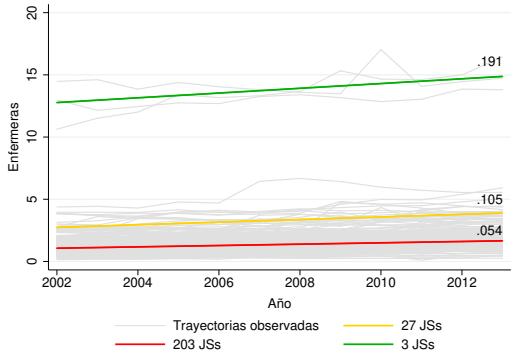
(d) Diferencia 2013-2001 según niveles de 2001



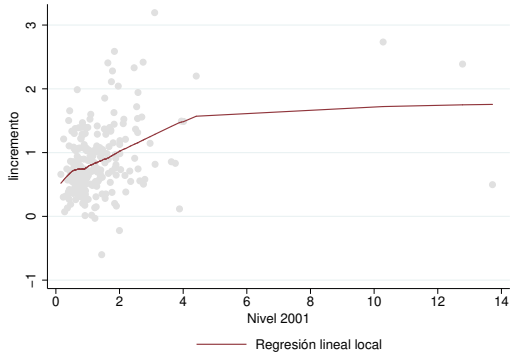
(continúa)

Enfermeras en contacto con pacientes

(e) Clases de trayectorias



(f) Diferencia 2013-2001 según niveles de 2001



Fuente: elaboración propia con datos de la Dirección General de Información en Salud.

La densidad se expresa por cada mil habitantes sin seguridad social asociada al trabajo: IMSS, ISSSTE, PEMEX, Defensa o Marina.

Los números en los descriptores de las gráficas a la izquierda hacen referencia al número de jurisdicciones sanitarias (JSs) pertenecientes a cada clase estimada.

Los números en la región interna de la gráficas a la izquierda se refieren a las pendientes estimadas para cada clase.

gación en el futuro inmediato. En particular, es posible acoplarle, por ejemplo, a otras técnicas de exploración por ponderación como la regresión ponderada geográficamente o de inferencia contrafactual como el análisis por descomposición. Ello permitiría incorporar la dimensión espacial al análisis observando dónde en un mapa son sistemáticamente diferentes las cantidades causales, o estimar la distribución contrafactual completa de una variable de interés.

Como principales resultados de nuestra investigación, ofrecemos pruebas estadísticas de que el SP ha tenido mayores efectos en las zonas geográficas del país menos vulnerables. Es usual que programas de desarrollo social tengan resultados no esperados e incluso indeseables. Este puede ser el caso del SP, que muy probablemente ha incrementado la desigualdad en la distribución de los recursos en salud en el país, dejando atrás a los segmentos de la población más vulnerables de acuerdo con el Índice de Marginación del Consejo Nacional de Población (CONAPO) (Fig. E.6).

Destacan así jurisdicciones sanitarias como la de Creel en el estado de Chihuahua y Motozintla en Chiapas, donde el SP exhibe sus menores efectos, en las que alrededor de 70 % de la población, vive en localidades que en 2010 eran consideradas de alta y muy alta marginación por el CONAPO; en contraste con jurisdicciones como Centro en el estado de Tabasco y Zacatecas del estado del mismo nombre, donde se estiman los mayores impactos del programa, en las que sólo alrededor del 4 % de su población vive en localidades de alta y muy alta marginación.

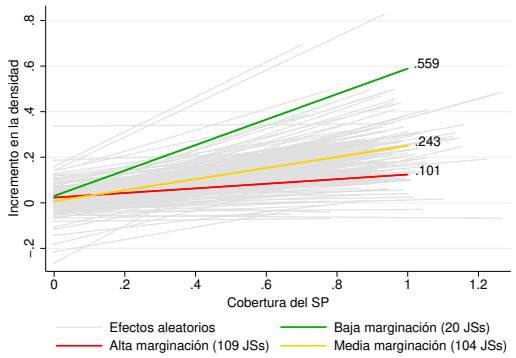
Nuestros resultados se encuentran en línea con lo observado recientemente por Cortés y Vargas [8] al analizar las trayectorias de marginación de los municipios mexicanos en el marco de la política nacional de desarrollo: el terreno ganado a la marginación ha venido acompañado de un aumento en la heterogeneidad.

Desde luego, no hay nada necesario en este patrón. Reducir las carencias no implica, en sentido lógico, ampliar la desigualdad. La contingencia de este fenómeno obliga a preguntarnos no sólo por la eficacia de la política pública de desarrollo social en general, sino por su eficiencia y la prioridad en los objetivos de la misma. Temas que reclaman

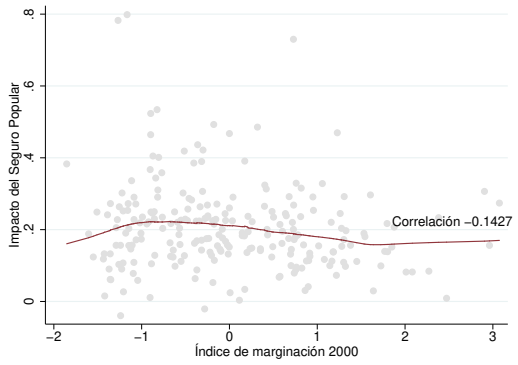
Figura E.6: Impacto del Seguro Popular sobre el incremento en la densidad de recursos de la Secretaría de Salud por jurisdicción sanitaria

Consultorios

(a) Clases de efectos



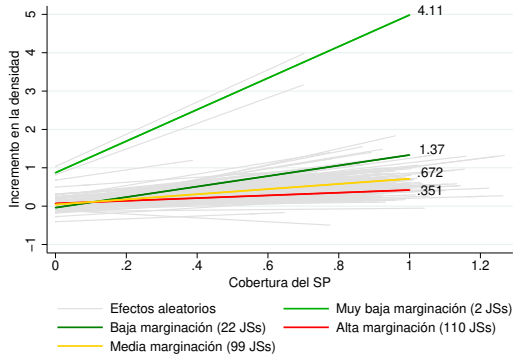
(b) Efectos por Índice de Marginación 2000



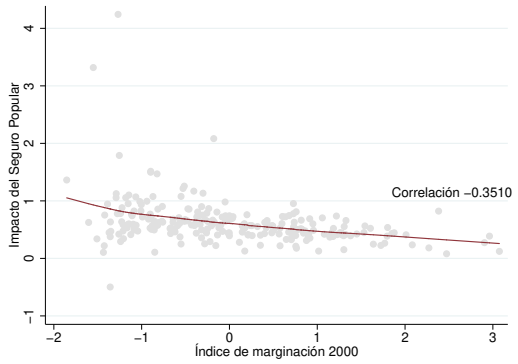
(continúa)

Médicos en contacto con pacientes

(c) Clases de efectos



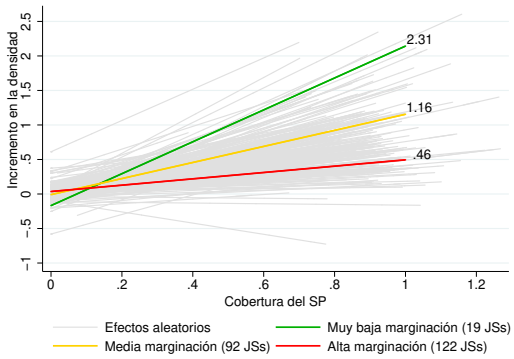
(d) Efectos por Índice de Marginación 2000



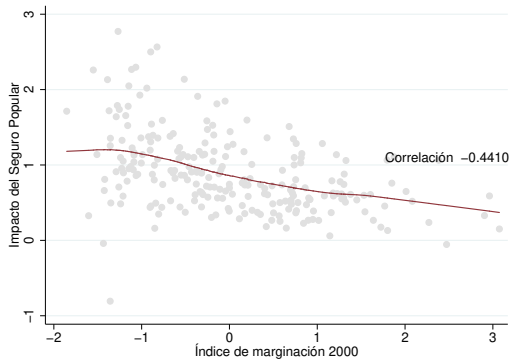
(continúa)

Enfermeras en contacto con pacientes

(e) Clases de efectos



(f) Efectos por Índice de Marginación 2000



Fuente: elaboración propia con datos de la Dirección General de Información en Salud y el Seguro Popular.

La densidad se expresa por cada mil habitantes sin seguridad social tradicional asociada al trabajo: IMSS, ISSSTE, PEMEX, Defensa o Marina. Sólo se cuentan aquellos médicos y enfermeras en labores clínicas (en contacto con pacientes).

La gráficas a la izquierda en cada subfigura presentan el resultado de estimar clases latentes de efectos aleatorios del Seguro Popular sobre el incremento en la densidad del respectivo recurso, ponderados por el inverso de la probabilidad de observar la respectiva historia de tratamiento. Los números en la región interna de de estas gráficas se refieren a las pendientes estimadas para cada clase. Las gráficas a la derecha presentan el indice de correlación de Pearson entre los mismos efectos aleatorios y el Índice de Marginación de las jurisdicciones sanitarias siguiendo el método de estimación del Consejo Nacional de Población http://www.conapo.gob.mx/es/CONAPO/Indices_de_Marginacion_Publicaciones.

Muy alta, alta, media, baja y muy baja marginación se refiere al nivel del Índice de Marginación promedio de las jurisdicciones sanitarias en cada clase.

particularmente la atención de contribuyentes y beneficiarios como primeros implicados.

Garantizar el derecho a la protección de la salud, como ideal regulativo, demanda de la sociedad evaluar la manera en que se invierten sus recursos escasos. ¿Cuál es la mejor senda a seguir para franquear el acceso efectivo a los servicios de salud en México? Los resultados de nuestra investigación contribuyen a esta discusión con información empírica sobre el probable impacto del SP sobre los recursos materiales y humanos de la SSA. Con ello abonamos a un conocimiento más preciso sobre el SPSS mexicano que permita mejorar el diseño de la política de salud en México. Esta es, sin duda, una de las cuestiones de mayor importancia en materia de política pública en nuestro país, cuya última expresión es el SP.

Nuestro argumento se presenta en tres capítulos, susceptibles de leerse por separado, escritos en inglés –lengua puente por excelencia en la disciplina– con la intención de alcanzar al mayor número de especialistas interesados. En el Capítulo 1 analizamos las implicaciones de diferentes supuestos causales en la evaluación del SP. Ahí mostramos la necesidad de extender la metodología de puntajes de propensión para analizar tratamientos continuos variables en el tiempo. En el Capítulo 2 presentamos un nuevo procedimiento semiparamétrico que permite inferir cantidades causales, asociadas a intervenciones dinámicas continuas. Esto último a través del maridaje de la bibliografía sobre tratamientos continuos con la de tratamientos dinámicos. El procedimiento propuesto permite estimar funciones dosis-respuesta, media y cuantil, a través de regresiones locales aplicadas a muestras ponderadas que controlan por variables confusoras variables en el tiempo. Es en este capítulo que se estima y discute el impacto del SP sobre la distribución de los recursos en salud. Dada la relativa complejidad detrás del procedimiento discutido en el Capítulo 2, en el Capítulo 3 presentamos un estimador robusto que facilita el análisis causal en contextos dinámicos con variables continuas. Por último se ofrecen algunos comentarios finales.

Cada vez más se reconoce que la política de desarrollo social en México se beneficiaría de un diseño más abierto, en un esquema de corresponsabilidad con la sociedad civil, que

promueva tanto la investigación académica como mecanismos de Contraloría Social. En último término, esperamos que la presente investigación ayude a estimular el surgimiento de un nuevo modo de hacer política pública.

Bibliography

[1] Javier Arias, Oliver Azuara, Pedro Bernal, James J. Heckman, and Cajeme Villarreal. Policies to promote growth and economic efficiency in mexico. Technical report, National Bureau of Economic Research, 2010.

[2] Leticia Ávila-Burgos, Edson Serván-Mori, Verónica J Wirtz, Sandra G Sosa-Rubí, and Aarón Salinas-Rodríguez. Efectos del Seguro Popular sobre el gasto en salud en hogares mexicanos a diez años de su implementación. *Salud Pública de México*, 55:S91–S99, 2013.

[3] Rodrigo Barros. Wealthier but not much Healthier: Effects of a Health Insurance Program for the Poor in Mexico. Technical report, Stanford University, 2008.

[4] Matthew Blackwell. A framework for dynamic causal inference in political science. *American Journal of Political Science*, 57(2):504–520, 2013.

[5] Mariano Bosch and Raymundo M. Campos-Vázquez. The Trade-offs of Social Assistance Programs in the Labor Market: The case of the "Seguro Popular" program in Mexico. Technical report, El Colegio de México, 2010.

[6] Raymundo M. Campos-Vazquez and Melissa A. Knox. Social protection programs and employment: the case of Mexico's Seguro Popular Program. *Economía Mexicana NUEVA ÉPOCA*, XXII, (2):403–448, 2013.

[7] Stephen R Cole and Miguel A Hernán. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6):656–664, 2008.

[8] Fernando Cortés and Delfino Vargas-Chanes. *Dos décadas de marginación en México: un efoque longitudinal*. Universidad Nacional Autónoma de México, 2016.

[9] Ángeles Cruz Martínez. El avance del Seguro Popular no ha sido igual en todas las entidades. *La Jornada*, 22 June:35, 2015. Available from: <http://www.jornada.unam.mx>. [28 July 2015].

[10] Alexis Diamond and Jasjeet S Sekhon. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3):932–945, 2013.

[11] Robert Duval Hernández and Ricardo Smith Ramírez. Informalidad y Seguro Popular bajo Mercados Laborales Segmentados. Technical report, Centro de Investigación y Docencia Económicas, A.C., 2010.

[12] Carlos A. Flores, Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C. Neumann. Estimating the effects of length of exposure to instruction in a training program: the case of job corps. *Review of Economics and Statistics*, 94(1):153–171, 2012.

[13] Christian Fong, Chad Hazlett, and Kosuke Imai. Parametric and Nonparametric Covariate Balancing Propensity Score for General Treatment Regimes. Mimeo, December 2015.

[14] Christian Fong, Marc Ratkovic, Chad Hazlett, and Kosuke Imai. Cbps: R package for covariate balancing propensity score. available at the Comprehensive R Archive Network (CRAN). 2015.

[15] Emmanuela Gakidou, Rafael Lozano, Eduardo González-Pier, et al. Evaluación del impacto de la Reforma Mexicana de Salud 2001-2006: un informe inicial. *Revista de Salud Publica*, 49 supl 1:S88–S109, 2007.

[16] John C Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338, 1966.

[17] John C Gower. A general coefficient of similarity and some of its properties. *Biometrics*, pages 857–871, 1971.

[18] Jeffrey Grogger, Tamara Arnold, Ana Sofía León, and Alejandro Ome. Efectos del Seguro Popular sobre el gasto en salud, utilización de servicios médicos, y nivel de salud, y resultados a largo plazo del experimento del Seguro Popular en México y evidencia de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). Technical report, Centro de Investigación y Docencia Económicas, A.C., 2011.

[19] Miguel A Hernán and Tyler J VanderWeele. Compound treatments and transportability of causal inference. *Epidemiology (Cambridge, Mass.)*, 22(3):368, 2011.

[20] Jinneth Hernández-Torres, Leticia Avila-Burgos, Atanacio Valencia-Mendoza, and Ofelia Poblano-Verástegui. Evaluación inicial del seguro popular sobre el gasto catastrófico en salud en México. *Revista de Salud Publica*, 10 (1):18–32, 2008.

[21] Keisuke Hirano and Guido W. Imbens. *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, volume 226164, chapter The propensity score with continuous treatments, pages 73–84. John Wiley & Sons, Ltd, 2005.

[22] Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[23] Curtis Huffman. Causal Inference with Time-Varying Continuous Interventions: Seguro Popular and the Supply of Health Care Services in Mexico. Technical report, El Colegio de México, Centro de Estudios Económicos, 2015.

[24] Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263, 2014.

[25] Kosuke Imai and Marc Ratkovic. Robust Estimation of Inverse Probability Weights for Marginal Structural Models. *Journal of the American Statistical Association*, 110(511):1013–1023, 2015.

[26] Kosuke Imai and David A Van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467):854–866, 2004.

[27] Guido W Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000.

[28] Gary King, Emmanuela Gakidou, Kosuke Imai, et al. Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. *Lancet*, 373(9673):1447–54, 2009.

[29] Gary King and Langche Zeng. The dangers of extreme counterfactuals. *Political Analysis*, 14(2):131–159, 2006.

[30] Gary King and Langche Zeng. When can history be our guide? the pitfalls of counterfactual inference. *International Studies Quarterly*, 51(1):183–210, 2007.

[31] Felicia Marie Knaul, Eduardo González-Pier, Octavio Gómez-Dantés, et al. The quest for universal health coverage: achieving social protection for all in Mexico. *The Lancet*, 380(9849):1259–1279, 2012.

[32] Melissa Knox. Health Insurance for All: An Evaluation of Mexico's Seguro Popular Program. *Unpublished manuscript, Department of Economics, University of California Berkley*, 2008.

[33] Roger Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.

[34] Serge-Christophe Kolm. Unequal inequalities. I. *Journal of Economic Theory*, 12(3):416–442, 1976.

[35] Serge-Christophe Kolm. Unequal inequalities. II. *Journal of Economic Theory*, 13(1):82–111, 1976.

[36] Ioana Marinescu and Oliver Azuara. La informalidad del empleo y la expansión de los sistemas de protección social. El caso de México. Technical report, Centro de Investigación y Docencia Económicas, A.C., 2010.

[37] Susan Parker and Luís N. Ruvalcaba. Identificación y análisis de los efectos en las condiciones de salud de los afiliados al Seguro Popular. Technical report, Centro de Investigación y Docencia Económicas, A.C., 2010.

[38] Judea Pearl. *Causality*. Cambridge university press, 2009.

[39] Nigel Rice and Peter C Smith. Ethics and geographical equity in health care. *Journal of Medical Ethics*, 27(4):256–261, 2001.

[40] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.

[41] James M Robins. *Statistical models in epidemiology, the environment, and clinical trials*, chapter Marginal structural models versus structural nested models as tools for causal inference, pages 95–133. Springer, 2000.

[42] James M Robins and Miguel A Hernán. *Longitudinal data analysis*, chapter Estimation of the causal effects of time-varying exposures, pages 553–599. CRC Press, 2008.

[43] James M Robins, Miguel Angel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[44] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[45] Paul. R. Rosenbaum and Donald B. Rubin. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524, 1984.

[46] Donald B Rubin. Matching to remove bias in observational studies. *Biometrics*, 29(1):159–183, 1973.

[47] Donald B. Rubin and Neal Thomas. Characterizing the effect of matching using linear propensity score methods with normal distributions. *Biometrika*, 79(4):797–809, 1992.

[48] Sistema de Protección Social en Salud. Informe de Resultados. Technical report, Secretaría de Salud, 2014.

[49] Jerzy Splawa-Neyman, DM Dabrowska, TP Speed, et al. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5(4):465–472, 1990.

[50] Graciela Teruel Belismelis, Manuel Castro, and Roberto Guadarrama. Estudio sobre los efectos del Seguro Popular en la utilización de servicios médicos y en la salud de los afiliados. Technical report, Centro de Investigación y Docencia Económicas, A.C., 2012.

[51] Tyler J VanderWeele and Miguel A Hernan. Causal inference under multiple versions of treatment. *Journal of causal inference*, 1(1):1–20, 2013.

[52] Shandong Zhao, David A. van Dyk, and Kosuke Imai. Propensity-Score Based Methods for Causal Inference in Observational Studies with Fixed Non-Binary Treatments. Technical report, University of California, Irvine, 2013.

Appendix A

Tables and Derivations

A.1 Tables from Chapter 1

Table A.1: Descriptive Statistics of the Sanitary Jurisdictions

Variable[a]	Sample	Mean	Std. Dev.	Min	Max	Obs.
SP coverage[b]	Full[c]	0.56	0.19	0.06	1.08	233
	Trimmed[d]	0.55	0.17	0.13	0.94	194
Doctor's offices	Full	0.54	0.35	0.17	3.05	233
	Trimmed	0.45	0.20	0.17	1.53	194
Staffed hospital beds	Full	0.66	1.01	0.00	9.70	233
	Trimmed	0.42	0.30	0.00	2.16	194
Non-staffed hospital beds	Full	0.76	0.49	0.00	3.58	233
	Trimmed	0.67	0.37	0.00	1.69	194
Physicians with day-to-day contact with patients	Full	1.06	0.98	0.22	10.47	233
	Trimmed	0.81	0.36	0.22	2.21	194
Physicians without day-to-day contact with patients	Full	0.09	0.16	0.00	1.63	233
	Trimmed	0.06	0.05	0.00	0.24	194

Variable[a]	Sample	Mean	Std. Dev.	Min	Max	Obs.
Nurses with day-to-day contact	Full	1.39	1.47	0.16	13.72	233
with patients	Trimmed	1.02	0.49	0.16	3.12	194
Nurses without day-to-day	Full	0.11	0.16	0.00	1.34	233
contact with patients	Trimmed	0.07	0.07	0.00	0.53	194

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records

[a] Other than SP coverage, variables are expressed per thousand population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

[b] Affiliates as proportion of the population without traditional health insurance.

[c]The full sample result from aggregating the municipality level data to the sanitary jurisdiction level.

[d]The trimmed sample results from dropping the farthest away observations in the covariate space according to Gower's metric.

Table A.2: Average Treatment Effects for Dichotomized SP coverage on Sanitary Jurisdictions[a]

Result Variable 2010-2001			Bootstrap			Normal-based	
(Change in Density)[b]	Sample	Coef.	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Doctor's offices	UM[c]	0.06	0.02	2.94	0.00	0.02	0.10
	CBPS[d]	0.05	0.02	2.88	0.00	0.02	0.08
	GM[e]	0.05	0.02	3.05	0.00	0.02	0.08
	NP-CBPS[f]	0.05	0.02	2.50	0.01	0.01	0.08
Physicians with day-to-day contact with patients	UM	0.12	0.04	3.38	0.00	0.05	0.19
	CBPS	0.13	0.04	3.34	0.00	0.05	0.20
	GM	0.11	0.03	3.23	0.00	0.04	0.18
	NP-CBPS	0.13	0.05	2.91	0.00	0.04	0.23
Physicians without day-to-day contact with patients	UM	0.01	0.01	1.40	0.16	0.00	0.02
	CBPS	0.00	0.01	0.92	0.36	-0.01	0.02
	GM	0.00	0.01	0.11	0.91	-0.01	0.01
	NP-CBPS	0.00	0.01	0.77	0.44	-0.01	0.02

Table continues

Result Variable 2010-2001			Bootstrap			Normal-based	
(Change in Density)[b]	Sample	Coef.	Std. Err.	z	$p > z$	[95% Conf.	Interval]
	UM	0.15	0.05	3.15	0.00	0.06	0.24
Nurses with day-to-day	CBPS	0.18	0.05	3.75	0.00	0.09	0.28
contact with patients	GM	0.13	0.05	2.83	0.01	0.04	0.23
	NP-CBPS	0.20	0.06	3.53	0.00	0.09	0.31
	UM	0.02	0.01	2.51	0.01	0.00	0.04
Nurses without day-to-day	CBPS	0.02	0.01	2.78	0.01	0.01	0.03
contact with patients	GM	0.01	0.01	1.61	0.11	0.00	0.02
	NP-CBPS	0.01	0.01	2.29	0.02	0.00	0.02

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records
Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] Treatment variable results from dichotomizing Seguro Popular coverage in 2009 around its median.

[b] Per thousand population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

[c] Unmatched sample refers to the 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data.

[d] Sample matched using Genetic Matching.

[e] Sample matched using Covariate Balance Propensity Score.

[f] Sample matched using Non-parametric Covariate Balance Propensity Score.

Table A.3: Covariate-Balance for SP coverage with Balanced Samples with respect to Dichotomized Treatment on Sanitary Jurisdictions[a]

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Doctor's offices	Dichotomous	UM	0.10	0.03	3.61	0.00	0.05	0.16
		CBPS	0.03	0.02	1.91	0.06	0.00	0.06
		GM	0.01	0.01	1.06	0.29	-0.01	0.03
		NP-CBPS	0.00	0.00	0.24	0.81	0.00	0.00
	SP coverage	UM	0.44	0.07	5.94	0.00	0.29	0.58
		CBPS	0.30	0.07	4.37	0.00	0.16	0.43
		GM	0.23	0.05	4.27	0.00	0.13	0.34
		NP-CBPS	0.24	0.07	3.31	0.00	0.10	0.38

Table continues

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Staffed hospital beds	Dichotomous	UM	-0.05	0.04	-1.20	0.23	-0.14	0.03
		CBPS	-0.01	0.02	-0.25	0.80	-0.05	0.04
		GM	-0.02	0.02	-0.97	0.33	-0.06	0.02
		NP-CBPS	0.00	0.00	1.22	0.22	0.00	0.00
	SP coverage	UM	-0.17	0.13	-1.29	0.20	-0.42	0.09
		CBPS	-0.06	0.10	-0.58	0.56	-0.25	0.14
		GM	-0.06	0.10	-0.55	0.58	-0.26	0.14
		NP-CBPS	-0.01	0.09	-0.07	0.94	-0.18	0.16

Table continues

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Non-staffed hospital beds	Dichotomous	UM	0.27	0.05	5.46	0.00	0.17	0.36
		CBPS	0.05	0.03	1.79	0.07	0.00	0.10
		GM	0.02	0.02	1.31	0.19	-0.01	0.06
		NP-CBPS	0.00	0.00	-0.17	0.86	-0.01	0.00
	SP coverage	UM	1.01	0.14	7.47	0.00	0.75	1.28
		CBPS	0.60	0.14	4.20	0.00	0.32	0.88
		GM	0.51	0.12	4.29	0.00	0.27	0.74
		NP-CBPS	0.51	0.16	3.14	0.00	0.19	0.83

Table continues

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Physicians with day-to-day contact with patients	Dichotomous	UM	0.12	0.05	2.38	0.02	0.02	0.22
		CBPS	0.05	0.03	1.59	0.11	-0.01	0.10
		GM	0.01	0.01	0.61	0.55	-0.02	0.04
		NP-CBPS	0.00	0.00	0.30	0.77	0.00	0.00
	SP coverage	UM	0.56	0.13	4.13	0.00	0.29	0.82
		CBPS	0.40	0.11	3.68	0.00	0.19	0.61
		GM	0.29	0.09	3.23	0.00	0.11	0.47
		NP-CBPS	0.31	0.10	3.27	0.00	0.12	0.50

Table continues

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Physicians without day-to-day contact with patients	Dichotomous	UM	-0.01	0.01	-0.91	0.36	-0.02	0.01
		CBPS	0.00	0.00	-0.08	0.94	-0.01	0.01
		GM	0.00	0.00	-0.83	0.41	-0.01	0.00
		NP-CBPS	0.00	0.00	0.83	0.41	0.00	0.00
	SP coverage	UM	-0.03	0.02	-1.12	0.26	-0.07	0.02
		CBPS	-0.01	0.02	-0.54	0.59	-0.05	0.03
		GM	0.00	0.02	0.01	1.00	-0.04	0.04
		NP-CBPS	0.00	0.02	-0.01	1.00	-0.04	0.04

Table continues

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Nurses with day-to-day contact with patients	Dichotomous	UM	0.10	0.07	1.38	0.17	-0.04	0.24
		CBPS	0.06	0.04	1.50	0.13	-0.02	0.14
		GM	0.01	0.02	0.23	0.82	-0.04	0.05
		NP-CBPS	0.00	0.00	0.76	0.45	0.00	0.00
	SP coverage	UM	0.51	0.18	2.78	0.01	0.15	0.86
		CBPS	0.41	0.15	2.85	0.00	0.13	0.70
		GM	0.33	0.13	2.54	0.01	0.08	0.59
		NP-CBPS	0.32	0.13	2.44	0.02	0.06	0.58

Table continues

				Bootstrap			Normal-based	
Variable 2001[b]	Treatment	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Nurses without day-to-day contact with patients	Dichotomous	UM	-0.02	0.01	-1.63	0.10	-0.04	0.00
		CBPS	-0.01	0.01	-1.10	0.27	-0.02	0.00
		GM	-0.01	0.00	-1.43	0.15	-0.01	0.00
		NP-CBPS	0.00	0.00	0.37	0.71	0.00	0.00
	SP coverage	UM	-0.06	0.03	-1.75	0.08	-0.12	0.01
		CBPS	-0.02	0.03	-0.94	0.35	-0.07	0.03
		GM	-0.02	0.02	-0.85	0.40	-0.06	0.03
		NP-CBPS	0.00	0.02	0.01	0.99	-0.04	0.04

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] SP coverage refers to affiliates in 2009 as proportion of the population without traditional (work-related) health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy). Dichotomous treatment variable results from dichotomizing Seguro Popular coverage in 2009 around its median.

[b] Per thousand population without traditional health insurance.

[c] UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, CBPS to the sample matched using Covariate Balance Propensity Score, GM to the sample matched using Genetic Matching and NP-CBPS to the sample matched using Non-parametric Covariate Balance Propensity Score, all with respect to dichotomized treatment.

[d] Coefficients result from regressing each covariate in 2001 on the dichotomized and original SP coverage in 2009.

Table A.4: Covariate-Balance for SP coverage on Sanitary Jurisdictions[a]

Variable 2001[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
	UM	0.44	0.08	5.83	0.00	0.29	0.59
	OLS-SW	-0.06	0.17	-0.35	0.73	-0.39	0.27
Doctor's offices	CBPS-OVER	0.17	0.16	1.06	0.29	-0.14	0.48
	CBPS-EXACT	0.03	0.06	0.50	0.62	-0.09	0.15
	NP-CBPS	0.00	0.02	-0.02	0.99	-0.03	0.03
	UM	-0.17	0.13	-1.25	0.21	-0.42	0.09
	OLS-SW	-0.03	0.17	-0.16	0.88	-0.37	0.31
Staffed hospital beds	CBPS-OVER	0.09	0.15	0.65	0.52	-0.19	0.38
	CBPS-EXACT	0.00	0.05	0.00	1.00	-0.11	0.11
	NP-CBPS	0.00	0.01	0.09	0.93	-0.03	0.03

Table continues

			Bootstrap			Normal-based	
Variable 2001[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Non-staffed hospital beds	UM	1.01	0.14	7.16	0.00	0.74	1.29
	OLS-SW	-0.08	0.35	-0.22	0.83	-0.75	0.60
	CBPS-OVER	0.44	0.32	1.37	0.17	-0.19	1.06
	CBPS-EXACT	0.05	0.11	0.46	0.64	-0.16	0.26
	NP-CBPS	0.00	0.04	-0.02	0.99	-0.09	0.08
Physicians with day-to-day contact with patients	UM	0.56	0.13	4.28	0.00	0.30	0.81
	OLS-SW	-0.10	0.27	-0.38	0.70	-0.63	0.42
	CBPS-OVER	0.31	0.18	1.70	0.09	-0.05	0.68
	CBPS-EXACT	0.04	0.08	0.52	0.60	-0.11	0.19
	NP-CBPS	0.00	0.02	-0.02	0.99	-0.04	0.04

Table continues

			Bootstrap			Normal-based	
Variable 2001[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Physicians without day-to-day contact with patients	UM	-0.03	0.02	-1.13	0.26	-0.07	0.02
	OLS-SW	0.02	0.04	0.71	0.48	-0.04	0.09
	CBPS-OVER	0.02	0.03	0.73	0.46	-0.04	0.08
	CBPS-EXACT	0.00	0.01	0.42	0.67	-0.01	0.02
	NP-CBPS	0.00	0.00	0.06	0.96	0.00	0.00
Nurses with day-to-day contact with patients	UM	0.51	0.17	2.91	0.00	0.17	0.85
	OLS-SW	0.00	0.37	0.01	0.99	-0.72	0.73
	CBPS-OVER	0.38	0.23	1.68	0.09	-0.06	0.82
	CBPS-EXACT	0.05	0.09	0.52	0.60	-0.13	0.22
	NP-CBPS	0.00	0.02	0.05	0.96	-0.04	0.04

Table continues

			Bootstrap			Normal-based	
Variable 2001[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Nurses without day-to-day contact with patients	UM	-0.06	0.03	-1.74	0.08	-0.12	0.01
	OLS-SW	0.01	0.05	0.20	0.84	-0.09	0.12
	CBPS-OVER	0.03	0.05	0.59	0.55	-0.06	0.12
	CBPS-EXACT	0.00	0.01	-0.08	0.94	-0.02	0.02
	NP-CBPS	0.00	0.00	0.07	0.95	-0.01	0.01

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] SP coverage refers to affiliates in 2009 as proportion of the population without traditional (work-related) health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

[b] Per thousand population without traditional health insurance.

[c] UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance.

[d] Coefficients result from regressing each covariate in 2001 on the SP coverage in 2009.

Table A.5: Average Treatment Effect of SP coverage on Sanitary Jurisdictions[a]

Result Variable 2010-2001			Bootstrap			Normal-based	
(Change in Density)[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Doctor's offices	UM	0.27	0.06	4.74	0.00	0.16	0.38
	OLS-SW	0.25	0.08	3.01	0.00	0.09	0.41
	CBPS-OVER	0.20	0.06	3.18	0.00	0.08	0.32
	CBPS-EXACT	0.15	0.09	1.66	0.10	-0.03	0.32
	NP-CBPS	0.19	0.08	2.33	0.02	0.03	0.35
Physicians with day-to-day contact with patients	UM	0.49	0.10	4.72	0.00	0.28	0.69
	OLS-SW	0.68	0.22	3.06	0.00	0.24	1.11
	CBPS-OVER	0.69	0.13	5.28	0.00	0.43	0.94
	CBPS-EXACT	0.70	0.15	4.61	0.00	0.40	1.00
	NP-CBPS	0.64	0.17	3.71	0.00	0.30	0.97

Table continues

Result Variable 2010-2001			Bootstrap			Normal-based	
(Change in Density)[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Physicians without day-to-day contact with patients	UM	0.02	0.02	1.18	0.24	-0.02	0.06
	OLS-SW	-0.02	0.03	-0.82	0.41	-0.07	0.03
	CBPS-OVER	0.01	0.02	0.25	0.81	-0.04	0.05
	CBPS-EXACT	0.01	0.03	0.20	0.84	-0.05	0.06
	NP-CBPS	-0.01	0.03	-0.23	0.81	-0.06	0.04
Nurses with day-to-day contact with patients	UM	0.52	0.16	3.26	0.00	0.21	0.83
	OLS-SW	0.87	0.27	3.23	0.00	0.34	1.40
	CBPS-OVER	0.85	0.17	4.89	0.00	0.51	1.19
	CBPS-EXACT	0.93	0.23	4.10	0.00	0.49	1.38
	NP-CBPS	0.79	0.24	3.22	0.00	0.31	1.27

Table continues

Result Variable 2010-2001			Bootstrap			Normal-based	
(Change in Density)[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Nurses without day-to-day contact with patients	UM	0.08	0.03	2.71	0.01	0.02	0.14
	OLS-SW	0.04	0.03	1.35	0.18	-0.02	0.11
	CBPS-OVER	0.06	0.03	2.03	0.04	0.00	0.12
	CBPS-EXACT	0.07	0.03	2.14	0.03	0.01	0.13
	NP-CBPS	0.05	0.03	1.89	0.06	0.00	0.11

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] SP coverage refers to affiliates in 2009 as proportion of the population without traditional health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

[b] Per thousand population without traditional health insurance.

[c] UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance.

[d] Coefficients result from regressing each result variable on the SP coverage in 2009.

Table A.6: Quantile Treatment Effect of SP coverage on Sanitary Jurisdictions[a]

/Sample[b]		UM		OLS-SW		CBPS-OVER		CBPS-EXACT		NP-CBPS	
Res. Variable[c]	Quantile	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$
	.10	0.15	0.07	0.02	0.84	0.16	0.09	-0.02	0.89	-0.02	0.86
	.25	0.16	0.00	0.15	0.01	0.14	0.00	0.15	0.06	0.15	0.02
Doctor's	.50	0.26	0.00	0.34	0.00	0.21	0.00	0.26	0.01	0.21	0.07
offices	.75	0.30	0.00	0.37	0.02	0.22	0.00	0.25	0.03	0.28	0.05
	.90	0.38	0.00	0.39	0.02	0.31	0.00	0.34	0.02	0.36	0.01
	.95	0.45	0.00	0.48	0.00	0.44	0.00	0.44	0.01	0.44	0.00

Table continues

/Sample[b]		UM		OLS-SW		CBPS-OVER		CBPS-EXACT		NP-CBPS	
Res. Variable[c]	Quantile	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$
Physicians with day-to-day contact with patients	.10	0.40	0.00	0.35	0.13	0.40	0.02	0.32	0.24	0.32	0.15
	.25	0.43	0.00	0.46	0.14	0.55	0.00	0.53	0.03	0.43	0.07
	.50	0.58	0.00	0.87	0.00	0.78	0.00	0.76	0.00	0.73	0.00
	.75	0.50	0.00	0.76	0.02	0.79	0.00	0.81	0.00	0.76	0.00
	.90	0.32	0.25	0.50	0.14	0.50	0.03	0.53	0.05	0.46	0.05
	.95	0.00	0.99	0.60	0.23	0.15	0.77	0.30	0.53	0.21	0.74

Table continues

/Sample[b]		UM		OLS-SW		CBPS-OVER		CBPS-EXACT		NP-CBPS	
Res. Variable[c]	Quantile	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$	Coef.[d]	$p > z$
	.10	0.42	0.02	0.28	0.26	0.47	0.03	0.41	0.14	0.39	0.12
Nurses with	.25	0.33	0.01	0.52	0.06	0.85	0.00	0.74	0.01	0.64	0.02
day-to-day	.50	0.57	0.01	1.21	0.00	1.23	0.00	1.36	0.00	1.15	0.00
contact with	.75	0.73	0.00	1.20	0.06	0.86	0.00	1.02	0.01	0.94	0.04
patients	.90	0.66	0.09	1.00	0.15	0.68	0.03	1.10	0.02	0.84	0.07
	.95	0.30	0.66	0.82	0.16	0.52	0.28	1.12	0.02	0.39	0.47

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] Affiliates in 2009 as proportion of the population without traditional (labor-related) health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

[b] UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance.

[c] 2010-2001 change in density (per thousand population without traditional health insurance).

[d] Coefficients result from a quantile regression of each result variable on the SP coverage in 2009.

Table A.7: Interaction Treatment Effect of SP coverage with Pre-treatment Covariates on Sanitary Jurisdictions[a]

/Sample[b]		UM		OLS-SW		CBPS-OVER		CBPS-EXACT		NP-CBPS	
Res. Var.[c]	Terms[d]	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$
Doctor's offices	SP	0.31	0.01	0.33	0.00	0.26	0.02	0.37	0.01	0.19	0.16
	SPxDO	-0.16	0.53	-0.38	0.21	-0.13	0.67	0.33	0.35	-0.26	0.37
	SPxSHB	-0.16	0.13	-0.05	0.72	-0.13	0.25	0.06	0.69	-0.03	0.84
	SPxNSHB	0.03	0.71	0.08	0.40	0.05	0.53	0.09	0.30	0.08	0.41
	SPxPWC	0.20	0.35	0.11	0.66	0.10	0.70	-0.60	0.09	0.28	0.27
	SPxPWOC	-0.65	0.21	-0.10	0.88	-0.18	0.73	0.01	0.99	-0.01	0.99
	SPxNWC	-0.09	0.38	-0.03	0.77	-0.07	0.52	-0.01	0.95	-0.17	0.17
	SPxNWOC	0.49	0.54	0.18	0.77	0.30	0.69	0.48	0.60	0.27	0.72

Table continues

/Sample[b]		UM		OLS-SW		CBPS-OVER		CBPS-EXACT		NP-CBPS	
Res. Var.[c]	Terms[d]	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$
	SP	0.53	0.00	0.69	0.01	0.64	0.00	0.74	0.00	0.52	0.05
Physicians	SPxDO	-0.55	0.20	-0.87	0.10	-0.47	0.34	0.01	0.99	-0.76	0.18
with	SPxSHB	0.10	0.66	0.10	0.68	0.10	0.68	0.32	0.27	0.20	0.51
day-to-day	SPxNSHB	-0.04	0.74	-0.05	0.75	-0.12	0.31	-0.06	0.74	0.02	0.90
contact	SPxPWC	0.24	0.46	0.20	0.64	0.21	0.57	-0.34	0.47	0.44	0.32
with	SPxPWOC	1.72	0.21	2.76	0.11	2.33	0.10	2.52	0.05	3.07	0.05
patients	SPxNWC	0.05	0.77	0.10	0.58	0.04	0.81	0.03	0.91	-0.11	0.63
	SPxNWOC	-1.13	0.11	-0.98	0.34	-1.15	0.13	-0.96	0.35	-1.48	0.18

Table continues

/Sample[b]		UM		OLS-SW		CBPS-OVER		CBPS-EXACT		NP-CBPS	
Res. Var.[c]	Terms[d]	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$	Coef.[e]	$p > z$
	SP	0.61	0.01	0.88	0.01	0.87	0.00	0.79	0.02	0.69	0.06
Nurses	SPxDO	-0.64	0.09	-0.79	0.09	-0.69	0.06	-0.85	0.08	-0.99	0.03
with	SPxSHB	0.55	0.05	0.25	0.44	0.48	0.12	0.49	0.16	0.46	0.21
day-to-day	SPxNSHB	-0.01	0.92	-0.12	0.54	-0.15	0.30	-0.12	0.56	0.03	0.89
contact	SPxPWC	0.40	0.26	0.35	0.46	0.55	0.15	0.93	0.06	0.77	0.11
with	SPxPWOC	2.25	0.10	2.56	0.24	1.84	0.18	2.07	0.22	3.11	0.04
patients	SPxNWC	-0.25	0.39	-0.07	0.83	-0.29	0.29	-0.43	0.23	-0.44	0.18
	SPxNWOC	-1.00	0.25	-0.57	0.66	-0.60	0.53	-0.40	0.75	-0.86	0.49

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] Affiliates in 2009 as proportion of the population without traditional (labor-related) health insurance: IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy.

[b] UM corresponds to the unmatched 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data, OLS-SW to the sample weighted using Robins' stabilized weights assuming Gaussian distribution fitted with all two-way interactions; CBPS-OVER to the sample weighted using Over-identified Covariate Balance Propensity Score, which gives equal importance to both correctly predict treatment-assignation and balancing covariates; CBPS-EXACT to the sample weighted using Exactly-identified Covariate Balance Propensity Score, which privileges covariate balance over the probabilistic model; and NPCBPS to the sample weighted using Non-parametric Covariate Balance Propensity Score, which dispenses with models altogether only minimizing covariate balance.

[c] 2010-2001 change in density (per thousand population without traditional health insurance).

[d] SP corresponds to Seguro Popular coverage, the rest of the terms correspond to the interaction of Seguro Popular coverage with pre-treatment covariates as follows: SPxDO to the interaction with the density of Doctor's Offices in 2001, SPxSHB to the interaction with the density of Staffed Hospital Beds, SPxNSHB to the interaction with the density of Non-Staffed Hospital Beds, SPxPWC to the interaction with the density of Physicians with day-to-day contact with patients, SPxPWoC to the interaction with the density of Physicians without day-to-day contact with patients, SPxNWC to the interaction with the density of Nurses with day-to-day contact with patients and SPxNOWC to the interaction with the density of Nurses without day-to-day contact with patients.

[e] Coefficients result from regressing each result variable on the SP coverage in 2009 and its interaction with all 7 pre-treatment covariates.

A.2 Derivation and Tables from Chapter 3

Equation 3.3 follows from plugging the Gaussian distribution,

$$
\begin{aligned}
SW_{it}\left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right) &= \prod_{t'=0}^{t} \frac{f_{\sigma^2}\left(A_{it'}\right)}{f_\theta\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} \\
&= \prod_{t'=0}^{t} \frac{\cancel{\frac{1}{\sigma\sqrt{2\pi}}} \exp\left\{-\frac{1}{2\sigma^2} A_t^2\right\}}{\cancel{\frac{1}{\sigma\sqrt{2\pi}}} \exp\left\{-\frac{1}{2\sigma^2}\left[A_t - \left(\underline{A}_{t-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_t^\top \boldsymbol{\beta}\right)\right]^2\right\}} \\
&= \prod_{t'=0}^{t} \exp\left\{\frac{1}{2\sigma^2}\left(\left[A_t - \left(\underline{A}_{t-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_t^\top \boldsymbol{\beta}\right)\right]^2 - A_t^2\right)\right\} \\
&= \prod_{t'=0}^{t} \exp\left\{\frac{1}{2\sigma^2}\left(\cancel{A_t^2} - 2A_t\left(\underline{A}_{t-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_t^\top \boldsymbol{\beta}\right) + \left(\underline{A}_{t-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_t^\top \boldsymbol{\beta}\right)^2 - \cancel{A_t^2}\right)\right\} \\
&= \prod_{t'=0}^{t} \exp\left[\frac{1}{2\sigma^2}\left\{-2A_{it'}\left(\underline{A}_{it'-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta}\right) + \left(\underline{A}_{it'-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta}\right)^2\right\}\right].
\end{aligned}
$$

Equation 3.4 follows from plugging the definition of Robins' stabilized weights,

$$
\begin{aligned}
&\mathrm{E}\left[SW_{it}\left(\underline{A}_{it}, \underline{\boldsymbol{X}}_{it}\right) \underline{A}_{it} \underline{\boldsymbol{X}}_{it}\right] \\
&\qquad = \mathrm{E}\left[\prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} \underline{A}_{it} \underline{\boldsymbol{X}}_{it}\right] \\
&= \int \left\{ \int \prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} \underline{A}_{it} dF\left(\underline{A}_{it} \mid \underline{\boldsymbol{X}}_{it}\right) \right\} \underline{\boldsymbol{X}}_{it} dF\left(\underline{\boldsymbol{X}}_{it}\right) \\
&\qquad = \mathrm{E}\left(\underline{A}_{\mathrm{it}}\right) \mathrm{E}\left(\underline{\boldsymbol{X}}_{\mathrm{it}}\right),
\end{aligned}
$$

where last equality follows from assuming that current variables can only be affected by

the past; and that the marginal distribution of A_t is independent across time periods, i.e.,

$$
\begin{aligned}
& f\left(\underline{A}_{it} \mid \underline{\boldsymbol{X}}_{it}\right) \\
& \quad = f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) f\left(A_{it-1} \mid \underline{A}_{it-2}, \underline{\boldsymbol{X}}_{it}\right) \cdots f\left(A_{i1} \mid \underline{\boldsymbol{X}}_{it}\right) \\
& \quad = f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) f\left(A_{it-1} \mid \underline{A}_{it-2}, \underline{\boldsymbol{X}}_{it-1}\right) \cdots f\left(A_{i1} \mid \boldsymbol{X}_{i1}\right) \\
& \quad = \prod_{t'=0}^{t} f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right).
\end{aligned}
$$

The elements of the last row and column of $\sum_{\boldsymbol{\theta}}\left(\underline{A}_t, \underline{\boldsymbol{X}}_t\right)$ are derived as follows.

$$
\begin{aligned}
\textstyle\sum \boldsymbol{\theta}\left[1,3\right] &= \textstyle\sum \boldsymbol{\theta}\left[3,1\right] \\
&= \mathrm{E}\left[g_1 g_3 \mid (\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\right] \\
&= \mathrm{E}\left[\frac{1}{\sigma^2}\left(A_{it} - \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right)(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}) \prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \mid (\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\right] \\
&= \frac{1}{\sigma^2}\left\{\int_{-\infty}^{\infty}\left(A_{it}^2 - A_{it}\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right)\prod_{t'=0}^{t}\frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) dA_{it}\right\}(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top} \\
&= \frac{1}{\sigma^2}\prod_{t'=0}^{t-1}\frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)}\left\{\int_{-\infty}^{\infty}\left(A_{it}^2 - A_{it}\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right)\frac{f\left(A_{it}\right)}{\cancel{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)}} \cancel{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)} dA_{it}\right\} \\
&\qquad (\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top} \\
&= \frac{1}{\sigma^2}\prod_{t'=0}^{t-1}\frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)}\left\{\cancel{\mathrm{E}\left(A_{it}^2\right)}^{\sigma^2} - \cancel{\mathrm{E}\left(A_{it}\right)}^{0}\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right\}(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top} \\
&= \prod_{t'=0}^{t-1}\frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)}(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}.
\end{aligned}
$$

Where the last equality follows from assuming the marginal distribution of A_t to be Gaussian with mean zero, and thus $\mathrm{E}\left(A_{it}\right) = 0$ and $\mathrm{E}\left(A_{it}^2\right) = \sigma^2$.

$$\sum \boldsymbol{\theta}[2,3] = \sum \boldsymbol{\theta}[3,2]$$

$$= \mathrm{E}\left[g_2 g_3 \mid (\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\right]$$

$$= \mathrm{E}\left[-\frac{1}{2\sigma^2}\left\{1 - \frac{1}{\sigma^2}\left[A_{it} - \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right]^2\right\} \prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \mid \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\right]$$

$$= \mathrm{E}\left[-\frac{1}{2\sigma^2}\left\{1 - \frac{1}{\sigma^2}\left[A_{it}^2 - 2A_{it}\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right) + \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)^2\right]\right\}\right.$$

$$\left. \prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \mid \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\right]$$

$$= \mathrm{E}\left[\frac{1}{2\sigma^4}\left\{A_{it}^2 - 2A_{it}\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right) + \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)^2 - \sigma^2\right\}\right.$$

$$\left. \prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \mid \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\right]$$

$$= \frac{1}{2\sigma^4} \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)}$$

$$\left\{\int_{-\infty}^{\infty}\left(A_{it}^3 - 2A_{it}^2\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right) + A_{it}\left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)^2 - A_{it}\sigma^2\right) \cancel{\frac{f\left(A_{it}\right)}{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)}} \cancel{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)} dA_{it}\right\}$$

$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)$$

$$= \frac{1}{2\sigma^4} \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} \left\{ \cancel{\mathrm{E}\left(A_{it}^3\right)}^{0} - 2\cancel{\mathrm{E}\left(A_{it}^2\right)}^{\sigma^2} \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta}\right) + \cancel{\mathrm{E}\left(A_{it}\right)}^{0} \left[\left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta}\right)^2 - \sigma^2 \right] \right\}$$

$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)$$

$$= \frac{1}{\cancel{2}\sigma^{\cancel{4}^{2}}} \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} \left\{ -\cancel{2}\sigma^{\cancel{2}} \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta}\right) \right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)$$

$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} \left\{ \frac{-\left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta}\right)}{\sigma^2} \right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right).$$

Where the last two equalities follow from assuming the marginal distribution of A_t to be Gaussian with mean zero, making its third moment equal zero, $\mathrm{E}\left(A_{it}^3\right)$, and its second moment equal its variance, $\mathrm{E}\left(A_{it}^2\right) = \sigma^2$.

$$\sum\nolimits_{\boldsymbol{\theta}}[3,3]$$

$$= \mathrm{E}\left[g_3 g_3 \mid (\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it})\right]$$

$$= \mathrm{E}\left[\prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \prod_{t'=0}^{t} \frac{f\left(A_{it'}\right)}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)} A_{it}\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \mid \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\right]$$

$$= \left\{\int_{-\infty}^{\infty} \prod_{t'=0}^{t} \frac{A_{it}^2 f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) dA_{it}\right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$

$$= \left\{\prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \int_{-\infty}^{\infty} \frac{A_{it}^2 f\left(A_{it}\right)^2}{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\cancel{2}}} \cancel{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)} dA_{it}\right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$

$$= \left\{\prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \int_{-\infty}^{\infty} \frac{A_{it}^2 f\left(A_{it}\right)^2}{f\left(A_{it} \mid \underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)} dA_{it}\right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$

$$= \left\{\prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \int_{-\infty}^{\infty} \frac{A_{it}^2 \left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[A_{it}\right]^2\right\}\right)^2}{\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[A_{it} - \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right]^2\right\}} dA_{it}\right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$

$$= \left\{\prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \int_{-\infty}^{\infty} \frac{A_{it}^2 \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{\cancel{2}} \left(\exp\left\{-\frac{1}{2\sigma^2}\left[A_{it}\right]^2\right\}\right)^{\cancel{2}}}{\cancel{\frac{1}{\sigma\sqrt{2\pi}}} \exp\left\{-\frac{1}{2\sigma^2}\left[A_{it} - \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right]^2\right\}} dA_{it}\right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$

$$= \left\{\prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \int_{-\infty}^{\infty} \frac{A_{it}^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{\sigma^2}\left[A_{it}\right]^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left[A_{it} - \left(\underline{A}_{it-1}^{\top}\boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top}\boldsymbol{\beta}\right)\right]^2\right\}} dA_{it}\right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$

$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \left\{ \int_{-\infty}^{\infty} \frac{A_{it}^2}{\sigma\sqrt{2\pi}} \exp\left(\frac{1}{2\sigma^2} \left(\left[A_{it} - \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta} \right) \right]^2 - 2\underline{A}_{it}^2 \right) \right) dA_{it} \right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)$$

$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^\top$$

$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \left\{ \int_{-\infty}^{\infty} \frac{A_{it}^2}{\sigma\sqrt{2\pi}} \exp\left(\frac{1}{2\sigma^2} \left(\cancel{A_{it}^2} - 2A_{it} \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta} \right) + \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta} \right)^2 - \cancel{2}\underline{A}_{it}^2 \right) \right) dA_{it} \right\}$$

$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^\top$$

$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \left\{ \int_{-\infty}^{\infty} \frac{A_{it}^2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \left(A_{it}^2 + 2A_{it} \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^\top \boldsymbol{\beta} \right) - \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta} \right)^2 \right) \right) dA_{it} \right\}$$

$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^\top$$

$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2}$$

$$\left\{ \int_{-\infty}^{\infty} \frac{A_{it}^2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \left(A_{it}^2 + 2A_{it} \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta} \right) + \left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta} \right)^2 - 2\left(\underline{A}_{it-1}^\top \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^\top \boldsymbol{\beta} \right)^2 \right) \right) dA_{it} \right\}$$

$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^\top$$

$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \left\{ \int_{-\infty}^{\infty} \frac{A_{it}^2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \left(\left[A_{it} + \left(\underline{A}_{it-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^{\top} \boldsymbol{\beta} \right) \right]^2 - 2 \left(\underline{A}_{it-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^{\top} \boldsymbol{\beta} \right)^2 \right) \right) dA_{it} \right\}$$
$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$
$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \left\{ \int_{-\infty}^{\infty} \frac{A_{it}^2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \left[A_{it} + \left(\underline{A}_{it-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top} \boldsymbol{\beta} \right) \right]^2 \right) dA_{it} \right\} \exp\left\{ \frac{\left(\underline{A}_{it-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top} \boldsymbol{\beta} \right)^2}{\sigma^2} \right\}$$
$$\left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}$$
$$= \prod_{t'=0}^{t-1} \frac{f\left(A_{it'}\right)^2}{f\left(A_{it'} \mid \underline{A}_{it'-1}, \underline{\boldsymbol{X}}_{it'}\right)^2} \left\{ \sigma^2 + \left(\underline{A}_{it-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top} \boldsymbol{\beta} \right)^2 \right\} \exp\left\{ \frac{\left(\underline{A}_{it-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it}^{\top} \boldsymbol{\beta} \right)^2}{\sigma^2} \right\} \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right) \left(\underline{A}_{it-1}, \underline{\boldsymbol{X}}_{it}\right)^{\top}.$$

Where the last equality follows from recognizing the Gaussian distribution of A_{it} with mean -$\left(\underline{A}_{it'-1}^{\top} \boldsymbol{\alpha} + \underline{\boldsymbol{X}}_{it'}^{\top} \boldsymbol{\beta} \right)$ and variance σ^2, whose second noncentral moment appears enclosed in the first set of brackets.

Table A.8: Covariate-Balance for Time-varying SP coverage on Sanitary Jurisdictions[a]

			Bootstrap			Normal-based	
Variable[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Doctor's offices	UW	0.10	0.03	3.75	0.00	0.05	0.15
	GLM	0.05	0.02	2.89	0.00	0.02	0.09
	CBIPW-OVER	0.01	0.06	0.20	0.84	-0.10	0.13
	CBIPW-EXACT	0.01	0.02	0.47	0.64	-0.03	0.05
Staffed hospital beds	UW	0.00	0.04	0.02	0.98	-0.07	0.07
	GLM	0.01	0.02	0.46	0.65	-0.03	0.05
	CBIPW-OVER	0.02	0.05	0.43	0.67	-0.07	0.11
	CBIPW-EXACT	0.01	0.02	0.40	0.69	-0.04	0.05
Non-staffed hospital beds	UW	0.24	0.07	3.61	0.00	0.11	0.37
	GLM	-0.01	0.03	-0.32	0.75	-0.06	0.04
	CBIPW-OVER	-0.08	0.12	-0.67	0.50	-0.31	0.15
	CBIPW-EXACT	0.01	0.02	0.40	0.69	-0.04	0.06

Table continues

			Bootstrap			Normal-based	
Variable[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Physicians with day-to-day contact with patients	UW	0.21	0.06	3.54	0.00	0.09	0.33
	GLM	0.13	0.04	3.66	0.00	0.06	0.21
	CBIPW-OVER	0.11	0.08	1.27	0.21	-0.06	0.27
	CBIPW-EXACT	0.02	0.04	0.55	0.58	-0.05	0.09
Physicians without day-to-day contact with patients	UW	-0.01	0.01	-1.37	0.17	-0.02	0.00
	GLM	0.00	0.00	-0.57	0.57	-0.01	0.01
	CBIPW-OVER	0.00	0.01	0.06	0.95	-0.03	0.03
	CBIPW-EXACT	0.00	0.00	0.25	0.80	-0.01	0.01
Nurses with day-to-day contact with patients	UW	0.25	0.09	2.88	0.00	0.08	0.41
	GLM	0.14	0.05	2.95	0.00	0.05	0.24
	CBIPW-OVER	0.11	0.13	0.82	0.41	-0.15	0.36
	CBIPW-EXACT	0.03	0.04	0.70	0.49	-0.06	0.12

Table continues

			Bootstrap			Normal-based	
Variable[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Nurses without day-to-day contact with patients	UW	-0.02	0.01	-1.93	0.05	-0.04	0.00
	GLM	0.00	0.01	-0.61	0.54	-0.01	0.01
	CBIPW-OVER	0.00	0.02	-0.12	0.91	-0.04	0.04
	CBIPW-EXACT	0.00	0.01	0.37	0.71	-0.01	0.01

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] SP coverage refers to affiliates as proportion of the population without traditional (work-related) health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

[b] Per thousand population without traditional health insurance.

[c] UW corresponds to the unweighted 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data; GLM to the sample weighted using OLS and an empty model as stabilizer; CBIPW-OVER to the sample weighted using over-identified correlation-breaking inverse-probability weights, which gives equal importance to both correctly predict treatment-assignation and balancing covariates and CBPS-EXACT to the sample weighted using Just-identified CBIPW, which privileges covariate balance over the probabilistic model.

[d] Coefficients result from a pooled regressions (2328 obs.) of each covariate on the yearly increment of SP coverage.

Table A.9: Average Treatment Effect of SP coverage on Sanitary Jurisdictions[a]

			Bootstrap			Normal-based	
Outcome Variable[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Doctor's offices	UW	0.20	0.01	16.90	0.00	0.18	0.23
	GLM	0.19	0.01	17.37	0.00	0.16	0.21
	CBIPW-OVER	0.19	0.04	5.31	0.00	0.12	0.26
	CBIPW-EXACT	0.19	0.01	16.03	0.00	0.17	0.22
Physicians with day-to-day contact with patients	UW	0.53	0.02	22.42	0.00	0.49	0.58
	GLM	0.51	0.02	21.50	0.00	0.46	0.56
	CBIPW-OVER	0.52	0.05	10.36	0.00	0.42	0.61
	CBIPW-EXACT	0.51	0.03	15.23	0.00	0.45	0.58

Table continues

			Bootstrap			Normal-based	
Outcome Variable[b]	Sample[c]	Coef.[d]	Std. Err.	z	$p > z$	[95% Conf.	Interval]
Nurses with day-to-day contact with patients	UW	0.73	0.04	19.00	0.00	0.66	0.81
	GLM	0.71	0.03	20.81	0.00	0.65	0.78
	CBIPW-OVER	0.72	0.05	15.46	0.00	0.63	0.82
	CBIPW-EXACT	0.72	0.04	18.09	0.00	0.64	0.80

Source: Own elaboration based on data from the Sistema Nacional de Información en Salud (National Health Information System) and Seguro Popular administrative records Estimates based on 1000 replications. Bootstrapping takes into account the matching algorithm.

[a] SP coverage refers to affiliates as proportion of the population without traditional (work-related) health insurance (IMSS, ISSSTE, Pemex, the Ministry of Defense or the Navy).

[b] Increment on the relative number (per thousand population without traditional health insurance) with respect to year prior the implementation of the program.

[c] UW corresponds to the unweighted 194 sanitary jurisdictions with the smallest mean distance to the rest of the observations in the data; GLM to the sample weighted using OLS and an empty model as stabilizer; CBIPW-OVER to the sample weighted using over-identified correlation-breaking inverse-probability weights, which gives equal importance to both correctly predict treatment-assignation and balancing covariates and CBPS-EXACT to the sample weighted using Just-identified CBIPW, which privileges covariate balance over the probabilistic model.

[d] Coefficients result from pooled regressions (2328 obs.) of each outcome variable on (cumulated) SP coverage.

Appendix B

Programming code (Stata®)

Do-file used to obtain the Correlation Breaking Inverse Probability of Treatment Weights in Chapter 3.

The next code is an adaptation of the code CBPSContinuous.r by Christian Fong, Chad Hazlett, and Kosuke Imai [14]. This code does not define a Stata® command (ado-file), readers are free to contact authors directly regarding any doubts or an electronic (.do) version.The authors would like to thank Christian Fong for useful comments. The usual disclaimer applies.

```
cap prog drop CBMSMContinuous
mata
mata clear
end

prog CBMSMContinuous, eclass
    version 13
    syntax varlist [if] [in] [, bal_only bal_only_init init_points(numlist) nelder_mead]
    marksample touse
    gettoken lhs rhs : varlist

    mat b = J(1,`:word count `rhs' _cons sigmasq',0)
    mat b_ort = J(1,`:word count `rhs' _cons sigmasq',0)
    summ `lhs' if `touse', d
    mat sw = J(r(N),1,0)
    mat sw_1 = J(r(N),1,0)
    mat balance = J(`:word count `rhs'', 1,0)
    mat basel = J(`:word count `rhs'', 1,0)
    sort id t

    tempvar bootrep bs_id
    tempname bs_mid
    bys id t: gen `bootrep'=_n
    sort id `bootrep' t
    egen bs_id = group(id `bootrep')


    mat colnames b = (intercept) `rhs' _sigmasq
    mat colnames b_ort = (intercept) `rhs' _sigmasq
    local init_points: subinstr local init_points " " ", ", all

    if ("`init_points'"!=""){
    mata: i_CBMSM_c("`lhs'", "`rhs'", "`touse'", "`bal_only'", "`bal_only_init'",
"`nelder_mead'", (`init_points'))
    }
    else {
    mata: i_CBMSM_c("`lhs'", "`rhs'", "`touse'", "`bal_only'", "`bal_only_init'",
"`nelder_mead'")
    }
    drop bs_id
    mat sw_transpose=sw'
    mat b_all=b_ort, sw_transpose
    eret post b_ort
    eret matrix sw sw
    eret matrix sw_1 sw_1
    eret matrix balance balance
    eret matrix basel basel
    eret local state "`=c(seed)'"
    eret di
end

mata:
    void i_lossf(todo,b,loss,g,H)
    {
        external Ttilde, Xtilde, info, time, stabilizers
        n = rows(Ttilde)
        k = cols(Xtilde)
        probs_min=1e-6
        cons = J(n,1,1)

        XtildeXtilde = cross(Xtilde,Xtilde)
        XtildeTtilde = cross(Xtilde,Ttilde)
        mcoef = pinv(XtildeXtilde)*XtildeTtilde
        e = Ttilde - (Xtilde)*mcoef
        sigmasq_mco=mean((Ttilde - Xtilde*mcoef):^2)
        Ttilde_hat=(Xtilde)*mcoef

        sigmasq = exp(log(sigmasq_mco)*b)
        probs_curr=normalden(Ttilde, Xtilde*(mcoef*b), sqrt(sigmasq))
        probs_curr=rowmin((cons:-probs_min,probs_curr))
```

```
69          probs_curr=rowmax((cons*probs_min,probs_curr))
70
71          probs_curr=log(probs_curr)
72
73          for (i=1; i<=rows(info); i++) {
74                          j0 = info[i, 1]
75                          j1 = info[i, 2]
76                          for (j=j0+1; j<=j1; j++) {
77                          probs_curr[j] = probs_curr[j-1] + probs_curr[j]
78                          }
79
80          }
81
82          w= stabilizers' - probs_curr
83          sw=w
84          swsq=2*sw
85          sw=exp(sw)
86          swsq=exp(swsq)
87
88          sw_1=J(rows(Ttilde),1,1)
89          for (i=1; i<=rows(info); i++) {
90                          j0 = info[i, 1]
91                          j1 = info[i, 2]
92                          for (j=j0+1; j<=j1; j++) {
93                          sw_1[j] = sw[j-1]
94                          }
95
96          }
97
98          swsq_1=J(rows(Ttilde),1,1)
99          for (i=1; i<=rows(info); i++) {
100                         j0 = info[i, 1]
101                         j1 = info[i, 2]
102                         for (j=j0+1; j<=j1; j++) {
103                         swsq_1[j] = swsq[j-1]
104                         }
105
106         }
107
108
109         w_curr=Ttilde:*sw
110
111         w_curr_del=1/n*Xtilde'*w_curr
112
113         score_1= 1/n*Xtilde'*(Ttilde-Xtilde*(mcoef*b))/sigmasq
114         score_2= 1/n*cons'*((Ttilde - Xtilde*(mcoef*b)):^2/(2*sigmasq^2) :- 1/(2*sigmasq))
115
116         gbar=(score_1 \score_2 \ w_curr_del)
117
118         n_identity_vec = J(rows(Ttilde),1,1)
119
120         Xtilde_1_1=1/sigmasq*Xtilde'*Xtilde
121         Xtilde_1_2=0*Xtilde'*n_identity_vec
122         Xtilde_1_3=Xtilde'*(Xtilde:*sw_1)
123         Xtilde_2_2=n_identity_vec'*n_identity_vec*(2*sigmasq^2)^(-1)
124         Xtilde_2_3=Xtilde'*((-Xtilde*(mcoef*b)/sigmasq):*sw_1)
125         Xtilde_3_3=Xtilde:*rowmin((exp((Xtilde*(mcoef*b)):^2/sigmasq + log(sigmasq :+ (
    Xtilde*(mcoef*b)):^2)), cons*10^250))
126         Xtilde_3_3=Xtilde'*(Xtilde_3_3:*swsq_1)
127
128         V=(1/n*(Xtilde_1_1,Xtilde_1_2,Xtilde_1_3) \ 1/n*(Xtilde_1_2',Xtilde_2_2,Xtilde_2_3')
     \ 1/n*(Xtilde_1_3,Xtilde_2_3,Xtilde_3_3))
129         invV=pinv(V)
130
131         loss1=gbar'*invV*gbar
132         loss=loss1*n
133
134     }
135
136
```

```
137
138
139
140       void i_loss(todo,b,loss,g,H)
141       {
142           external Ttilde, Xtilde, info, time, stabilizers
143
144           n = rows(Ttilde)
145           k = cols(Xtilde)
146           probs_min=1e-6
147           cons = J(n,1,1)
148
149           XtildeXtilde = cross(Xtilde,Xtilde)
150           XtildeTtilde = cross(Xtilde,Ttilde)
151
152           mcoef = pinv(XtildeXtilde)*XtildeTtilde
153           e = Ttilde - (Xtilde)*mcoef
154           sigmasq_mco=mean((Ttilde - Xtilde*mcoef):^2)
155
156           Ttilde_hat=(Xtilde)*mcoef
157
158           sigmasq = exp(b[cols(Xtilde)+1])
159
160           probs_curr=normalden(Ttilde, Xtilde*b[1..cols(Xtilde)]', sqrt(sigmasq))
161
162           probs_curr=rowmin((cons:-probs_min,probs_curr))
163           probs_curr=rowmax((cons*probs_min,probs_curr))
164
165           probs_curr=log(probs_curr)
166
167           for (i=1; i<=rows(info); i++) {
168                           j0 = info[i, 1]
169                           j1 = info[i, 2]
170                           for (j=j0+1; j<=j1; j++) {
171                           probs_curr[j] = probs_curr[j-1] + probs_curr[j]
172                           }
173
174           }
175
176           w= stabilizers' - probs_curr
177           sw=w
178           swsq=2*sw
179           sw=exp(sw)
180           swsq=exp(swsq)
181           sw_1=J(rows(Ttilde),1,1)
182           for (i=1; i<=rows(info); i++) {
183                           j0 = info[i, 1]
184                           j1 = info[i, 2]
185                           for (j=j0+1; j<=j1; j++) {
186                           sw_1[j] = sw[j-1]
187                           }
188
189           }
190
191           swsq_1=J(rows(Ttilde),1,1)
192           for (i=1; i<=rows(info); i++) {
193                           j0 = info[i, 1]
194                           j1 = info[i, 2]
195                           for (j=j0+1; j<=j1; j++) {
196                           swsq_1[j] = swsq[j-1]
197                           }
198           }
199
200
201           w_curr=Ttilde:*sw
202           w_curr_del=1/n*Xtilde'*w_curr
203
204           score_1= 1/n*Xtilde'*(Ttilde-Xtilde*b[1..cols(Xtilde)]')/sigmasq
205           score_2= 1/n*cons'*((Ttilde - Xtilde*b[1..cols(Xtilde)]'):^2/(2*sigmasq^2) :- 1/(2*
      sigmasq))
```

```
206
207          gbar=(score_1 \score_2 \ w_curr_del)
208
209          n_identity_vec = J(rows(Ttilde),1,1)
210
211          Xtilde_1_1=1/sigmasq*Xtilde'*Xtilde
212          Xtilde_1_2=0*Xtilde'*n_identity_vec
213          Xtilde_1_3=Xtilde'*(Xtilde:*sw_1)
214          Xtilde_2_2=n_identity_vec'*n_identity_vec*(2*sigmasq^2)^(-1)
215          Xtilde_2_3=Xtilde'*((-Xtilde*b[1..cols(Xtilde)]'/sigmasq):*sw_1)
216          Xtilde_3_3=Xtilde:*rowmin((exp((Xtilde*b[1..cols(Xtilde)]'):^2/sigmasq + log(sigmasq
      :+ (Xtilde*b[1..cols(Xtilde)]'):^2)), cons*10^250))
217          Xtilde_3_3=Xtilde'*(Xtilde_3_3:*swsq_1)
218
219          V=(1/n*(Xtilde_1_1,Xtilde_1_2,Xtilde_1_3) \ 1/n*(Xtilde_1_2',Xtilde_2_2,Xtilde_2_3')
      \ 1/n*(Xtilde_1_3,Xtilde_2_3,Xtilde_3_3))
220          invV=pinv(V)
221
222          loss1=gbar'*invV*gbar
223          loss=loss1*n
224
225          if (todo >= 1) {
226
227              dgbar_1_1 = quadcross(-Xtilde,Xtilde/sigmasq)
228              dgbar_1_2 = (-(Ttilde - Xtilde*b[1..cols(Xtilde)]')/(sigmasq^2))'*Xtilde
229              dgbar_2_2 = n_identity_vec'*(-(Ttilde - Xtilde*b[1..cols(Xtilde)]'):^2/(sigmasq^
      3) :+ (2*sigmasq^2)^(-1))
230
231              dgbar_3_1=(Xtilde):*(Ttilde-Xtilde*b[1..cols(Xtilde)]')/sigmasq
232
233
234              for (i=1; i<=rows(info); i++) {
235                          j0 = info[i, 1]
236                          j1 = info[i, 2]
237                         for (j=j0+1; j<=j1; j++) {
238                          dgbar_3_1[j,.] = dgbar_3_1[j-1,.] + dgbar_3_1[j,.]
239                          }
240
241              }
242
243              dgbar_3_1 = ((dgbar_3_1:*w_curr))'*Xtilde
244              dgbar_3_2 = (Ttilde - Xtilde*b[1..cols(Xtilde)]'):^2/(2*sigmasq^2) :- 1/(2*
      sigmasq)
245
246              for (i=1; i<=rows(info); i++) {
247                          j0 = info[i, 1]
248                          j1 = info[i, 2]
249                         for (j=j0+1; j<=j1; j++) {
250                          dgbar_3_2[j] = dgbar_3_2[j-1] + dgbar_3_2[j]
251                          }
252
253              }
254
255              dgbar_3_2 = (-w_curr:*(dgbar_3_2))'*Xtilde
256
257              dgbar = 1/n*((dgbar_1_1 \ dgbar_1_2*sigmasq),
258                           (dgbar_1_2'\ dgbar_2_2*sigmasq),
259                           (dgbar_3_1 \ dgbar_3_2*sigmasq))
260
261              gt = 2*n*dgbar*invV*gbar
262              g = gt'
263
264              }
265
266              if (todo == 2) {
267
268          H = 2*n*dgbar*invV*dgbar'
269          _makesymmetric(H)
270
271          }
```

```
272
273
274     }
275
276
277     void i_loss_bal_only(todo,b,loss,g,H)
278     {
279         external Ttilde, Xtilde, info, time, stabilizers
280
281         n = rows(Ttilde)
282         k = cols(Xtilde)
283         probs_min=1e-6
284         cons = J(n,1,1)
285
286         XtildeXtilde = cross(Xtilde,Xtilde)
287         XtildeTtilde = cross(Xtilde,Ttilde)
288         mcoef = pinv(XtildeXtilde)*XtildeTtilde
289         e = Ttilde - (Xtilde)*mcoef
290         sigmasq_mco=mean((Ttilde - Xtilde*mcoef):^2)
291         Ttilde_hat=(Xtilde)*mcoef
292
293         sigmasq = exp(b[cols(Xtilde)+1])
294
295         probs_curr=normalden(Ttilde, Xtilde*b[1..cols(Xtilde)]', sqrt(sigmasq))
296         /**/
297         probs_curr=rowmin((cons:-probs_min,probs_curr))
298         probs_curr=rowmax((cons*probs_min,probs_curr))
299
300         probs_curr=log(probs_curr)
301
302         for (i=1; i<=rows(info); i++) {
303                         j0 = info[i, 1]
304                         j1 = info[i, 2]
305                         for (j=j0+1; j<=j1; j++) {
306                         probs_curr[j] = probs_curr[j-1] + probs_curr[j]
307                         }
308
309         }
310
311         w= stabilizers' - probs_curr
312
313         sw=w
314         swsq=2*sw
315         sw=exp(sw)
316         swsq=exp(swsq)
317
318         sw_1=J(rows(Ttilde),1,1)
319         for (i=1; i<=rows(info); i++) {
320                         j0 = info[i, 1]
321                         j1 = info[i, 2]
322                         for (j=j0+1; j<=j1; j++) {
323                         sw_1[j] = sw[j-1]
324                         }
325
326         }
327
328         swsq_1=J(rows(Ttilde),1,1)
329         for (i=1; i<=rows(info); i++) {
330                         j0 = info[i, 1]
331                         j1 = info[i, 2]
332                         for (j=j0+1; j<=j1; j++) {
333                         swsq_1[j] = swsq[j-1]
334                         }
335
336         }
337
338
339         w_curr=Ttilde:*sw
340         w_curr_del=1/n*Xtilde'*w_curr
341         gbar=w_curr_del
```

```
342
343           invV=pinv(Xtilde'*Xtilde)
344
345           loss1=gbar'*invV*gbar
346           loss=loss1*n
347
348           if (todo >= 1) {
349
350
351               dgbar_3_1=(Xtilde):*(Ttilde-Xtilde*b[1..cols(Xtilde)]')/sigmasq
352
353
354               for (i=1; i<=rows(info); i++) {
355                           j0 = info[i, 1]
356                           j1 = info[i, 2]
357                          for (j=j0+1; j<=j1; j++) {
358                           dgbar_3_1[j,.] = dgbar_3_1[j-1,.] + dgbar_3_1[j,.]
359                           }
360
361               }
362
363
364               dgbar_3_1 = ((dgbar_3_1:*w_curr))'*Xtilde
365
366               dgbar_3_2 = (Ttilde - Xtilde*b[1..cols(Xtilde)]'):^2/(2*sigmasq^2) :- 1/(2*
    sigmasq)
367
368               for (i=1; i<=rows(info); i++) {
369                           j0 = info[i, 1]
370                           j1 = info[i, 2]
371                          for (j=j0+1; j<=j1; j++) {
372                           dgbar_3_2[j] = dgbar_3_2[j-1] + dgbar_3_2[j]
373                           }
374
375               }
376
377               dgbar_3_2 = (-w_curr:*(dgbar_3_2))'*Xtilde
378
379
380
381                   dgbar = 1/n*((dgbar_3_1) \
382                            sigmasq*(dgbar_3_2))
383
384                gt = 2*n*dgbar*invV*gbar
385                 g = gt'
386
387           }
388
389           if (todo == 2) {
390
391               H = 2*n*dgbar*invV*dgbar'
392               _makesymmetric(H)
393
394           }
395
396
397       }
398
399       void i_CBMSM_c(string scalar lhs, string scalar rhs, string scalar ok, string scalar
    bal_only, string scalar bal_only_init, string scalar NM,| real vector init_points)
400       {
401
402
403           external treat, X, Ttilde, Xtilde, panel_st, time, info, stabilizers
404           panel_st = st_data(., ("bs_id", "t"), ok)
405           time = st_data(., ("t"), ok)
406           info=panelsetup(panel_st, 1)
407           treat = st_data(., tokens(lhs), ok)
408           cons = J(rows(treat),1,1)
409           probs_min=1e-6
```

```
410
411           X = st_data(., tokens(rhs), ok)
412           X_orig=X
413           x_sd= (diagonal(sqrt(quadvariance(X))))'
414           X_mean=mean(X)
415           X=(X:-mean(X)):/ x_sd
416           X= cons, X
417           U = s = Vt = J(0,0,.)
418           svd(X,U,s,Vt)
419           X=U
420
421           Xtilde=(X :- mean(X))
422           T=(treat:-mean(treat))/(sqrt(quadvariance(treat)))
423           Ttilde=(T:-mean(T))
424
425           n = rows(Ttilde)
426           k = cols(Xtilde)
427
428           XtildeXtilde = cross(Xtilde,Xtilde)
429           XtildeTtilde = cross(Xtilde,Ttilde)
430           mcoef = pinv(XtildeXtilde)*XtildeTtilde
431           Ttilde_hat=(Xtilde)*mcoef
432           e = Ttilde - (Xtilde)*mcoef
433           sigmasq_mco=mean((Ttilde - Xtilde*mcoef):^2)
434
435             stabilizers=normalden(Ttilde[1,.], Ttilde_hat, sqrt(sigmasq_mco))
436
437             stabilizers=rowmin((cons:-probs_min,stabilizers))
438             stabilizers=rowmax((cons*probs_min,stabilizers))
439
440               i=2
441
442               while (i<=n) {
443                       probs=normalden(Ttilde[i,.], Ttilde_hat, sqrt(sigmasq_mco))
444                       /**/
445                       probs=rowmin((cons:-probs_min,probs))
446                       probs=rowmax((cons*probs_min,probs))
447
448                       stabilizers=(stabilizers,probs)
449                       i++
450               }
451
452               stabilizers=mean(stabilizers)
453
454               stabilizers=log(stabilizers)
455
456                 for (i=1; i<=rows(info); i++) {
457                                       j0 = info[i, 1]
458                                       j1 = info[i, 2]
459                                       for (j=j0+1; j<=j1; j++) {
460                                       stabilizers[j] = stabilizers[j-1] + stabilizers[j]
461                                       }
462
463                 }
464
465
466           if (args()==6) {
467               init = (mcoef', log(sigmasq_mco))
468
469           }
470
471           if (args()==7) {
472               init = init_points
473           }
474
475           S = optimize_init()
476
477           if (bal_only=="" & bal_only_init=="") {
478               optimize_init_evaluator(S, &i_loss())
479               }
```

```
480             else {
481             optimize_init_evaluator(S, &i_loss_bal_only())
482             }
483
484         optimize_init_which(S,"min")
485
486         if (NM=="") {
487             optimize_init_evaluatortype(S,"d2")
488             optimize_init_params(S,init)
489             optimize_init_technique(S, "nr dfp bfgs")
490             }
491             else {
492             optimize_init_evaluatortype(S,"d0")
493             optimize_init_params(S,init)
494             optimize_init_technique(S, "nm")
495
496             optimize_init_nmsimplexdeltas(S, J(1,cols(init),1/100000))
497
498             }
499
500         optimize_init_singularHmethod(S, "hybrid" )
501
502         optimize_init_trace_dots(S, "on")
503         optimize_init_trace_value(S, "on")
504         optimize_init_trace_tol(S, "on")
505         optimize_init_trace_step(S, "on")
506         optimize_init_trace_paramdiffs(S, "on")
507         optimize_init_trace_params(S, "on")
508         optimize_init_trace_gradient(S, "on")
509         optimize_init_trace_Hessian(S, "on")
510
511         _optimize(S)
512
513         if (_optimize(S)==0) {
514         p = optimize_result_params(S)
515         }
516
517         else {
518             optimize_init_evaluatortype(S,"d0")
519             optimize_init_params(S,init)
520             optimize_init_technique(S, "nm")
521             optimize_init_nmsimplexdeltas(S, J(1,cols(init),1))
522             p = optimize(S)
523             }
524
525         if (bal_only_init!="") {
526             init = p
527             if (bal_only=="")  {
528             optimize_init_evaluator(S, &i_loss())
529             }
530             else {
531             optimize_init_evaluator(S, &i_loss_bal_only())
532             }
533             optimize_init_evaluatortype(S,"d2")
534             optimize_init_params(S,init)
535             optimize_init_technique(S, "nr dfp bfgs")
536
537             _optimize(S)
538
539             if (_optimize(S)==0) {
540             p = optimize_result_params(S)
541             }
542             else {
543                 optimize_init_evaluatortype(S,"d0")
544                 optimize_init_params(S,init)
545                 optimize_init_technique(S, "nm")
546                 optimize_init_nmsimplexdeltas(S, J(1,cols(init),1))
547                 p = optimize(S)
548                 }
549
```

```
550             }
551
552
553           printf("Minimization report \n")
554           optimize_query(S)
555
556           mlf=optimize_result_value(S)
557           printf("Minimum loss function value = %9.4f \n", mlf)
558
559
560         beta_opt= p[1..cols(X)]
561         sigmasq_opt = exp(p[cols(X)+1])
562         probs_opt = normalden(Ttilde, Xtilde*beta_opt', sqrt(sigmasq_opt))
563         /**/
564         probs_opt = rowmin((cons:-probs_min,probs_opt))
565         probs_opt = rowmax((cons*probs_min,probs_opt))
566
567         probs_opt = log(probs_opt)
568
569         for (i=1; i<=rows(info); i++) {
570                               j0 = info[i, 1]
571                               j1 = info[i, 2]
572                               for (j=j0+1; j<=j1; j++) {
573                               probs_opt[j] = probs_opt[j-1] + probs_opt[j]
574                               }
575
576             }
577
578         w_opt = stabilizers' - probs_opt
579
580           sw_opt=w_opt
581           sw_opt=exp(sw_opt)
582
583           sw_opt_1=J(rows(Ttilde),1,1)
584           for (i=1; i<=rows(info); i++) {
585                           j0 = info[i, 1]
586                           j1 = info[i, 2]
587                           for (j=j0+1; j<=j1; j++) {
588                           sw_opt_1[j] = sw_opt[j-1]
589                           }
590           }
591
592          bal = J(cols(X_orig), 1,0)
593          baseline = J(cols(X_orig), 1,0)
594
595          for (j=1; j<=cols(X_orig); j++) {
596                               bal[j,1] = (mean(sw_opt:*X_orig[,j]:*treat) - mean(sw_opt:*X_orig[,j
      ])*mean(sw_opt:*treat)*n/sum(sw_opt))/(sqrt(mean(sw_opt:*X_orig[,j]:^2) - mean(sw_opt:*
      X_orig[,j]):^2*n/sum(sw_opt))*sqrt(mean(sw_opt:*treat:^2) - mean(sw_opt:*treat):^2*n/sum(
      sw_opt)))
597                           baseline[j,1] = quadcorrelation((treat, X_orig[,j]))[1,2]
598
599           }
600
601
602             bal
603             baseline
604
605           beta_opt = pinv(X'*X)*X'*(Xtilde*p[1..cols(X)]' :+ mean(T))
606           sinv = s':^-1
607           beta_opt = Vt'*diag(sinv)*beta_opt
608           beta_opt = beta_opt':/(1,x_sd)
609           beta_opt[.,1] = beta_opt[.,1]-(X_mean*beta_opt[2..cols(X)]')
610
611           st_replacematrix("b_ort",p)
612           st_replacematrix("b",(beta_opt,sigmasq_opt))
613           st_replacematrix("sw",sw_opt)
614           st_replacematrix("sw_1",sw_opt_1)
615           st_replacematrix("balance",bal)
616           st_replacematrix("basel",baseline)
```

```
617
618         }
619     end
620
```

Disclaimer

The software (do-file) is provided by the author as is and any express or implied warranties, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose are disclaimed. In no event shall the author be liable for any direct, indirect, incidental, special, exemplary, or consequential damages (including, but not limited to, procurement of substitute goods or services; loss of use, data, or profits; or business interruption) however caused and on any theory of liability, whether in contract, strict liability, or tort (including negligence or otherwise) arising in any way out of the use of this software, even if advised of the possibility of such damage.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of El Colegio de México is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it), and that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Ciudad de México, August 2016

Curtis Huffman Espinosa