Discussion Paper

Does Universal Long-term Care Insurance Boost Female Labor Force Participation? Macro-level Evidence

Michihito Ando, Masato Furuichi, Yoshihiro Kaneko

Discussion Paper E-2
April 8, 2019

Rikkyo Institute of Economic Research

Rikkyo University (St. Paul's University, Tokyo)
3-34-1 Nishi-Ikebukuro, Toshima-ku
Tokyo 171-8501, Japan

Does Universal Long-term Care Insurance Boost Female Labor Force Participation? Macro-level Evidence*

Michihito Ando† Masato Furuichi‡ Yoshihiro Kaneko §

April 8, 2019

Abstract

Exploiting the introduction of nationwide long-term care insurance (LTCI) in Japan and utilizing a synthetic control method, we examine how LTCI introduction has altered the trends of public expenditures on in-kind benefits for the elderly, public health expenditure, and female labor force participation. Estimation results using the panel data of OECD countries (1980s-2010s) suggest that LTCI introduction substantially increased in-kind benefits for the elderly, but we do not find any positive effect on female labor force participation for any age cohort. We also provide extended placebo tests based on permuted treatment assignment, resampled donor pools, and several test statistics to verify the above conclusions.

JEL classification: H42, H53, H61, I13, J21, J22

Keywords: long-term care insurance, synthetic control method, aggregate effect, female labor force participation

*We are grateful to Emile Cammeraat, Ayako Kondo, Sasaki Masaru, Fredrik Sävje, Reo Takaku, Shintaro Yamaguchi, and seminar participants at FISS, Keio University, Ritsumeikan University, University of Toyama, and IIPF for their comments and suggestions. This research is funded by a KAKENHI grants from the Japan Society for the Promotion of Science (grant number 15H03364, 17K03795, 18H00863). All errors are our own.

†Rikkyo University, Department of Economics, 3-34-1 Nishi-Ikebukuro,Toshima-ku, Tokyo Japan 171-8501, +81-3-3595-2984, michihito.ando@rikkyo.ac.jp.

‡Faculty of Economics, Teikyo University, Japan, masato@main.teikyo-u.ac.jp

§Faculty of Social Welfare, Japan College of Social Work, Japan, y-kaneko2@jcsw.ac.jp

1 Introduction

The introduction or expansion of a social insurance program is one of the most essential and controversial public policy issues in both developed and developing countries. A central topic of study and debate is public health insurance. The 2008 Medicaid expansion in Oregon and the Affordable Care Act in the U.S., for instance, are leading examples of policy changes that have induced numerous academic studies and public discussions (Obama 2016; Sommers et al. 2017).

While public long-term care (LTC) insurance is much less studied than public health insurance, LTC is important for both those who need care and their family members, in particular female informal caregivers, and the effects of informal caregiving on female labor supply have been intensively studied in health and labor economics.[1] Another important topic in the empirical literature on LTC in economics is the relationship between informal and formal LTC.[2]

On the other hand, only a few papers examine the effects of public LTC insurance (hereafter LTCI) on female labor supply (Shimizutani et al. 2008; Tamiya et al. 2011; Sugawara and Nakamura 2014; Fukahori et al. 2015; Geyer and Korfhage 2015; 2018; Fu et al. 2017). We can point out at least two reasons why the effects of public LTCI are not much studied. First, there are only a handful of developed countries (Luxemburg, the Netherlands, Germany, Japan, and South Korea) that have introduced independent public LTCI programs. Public LTC services in many other countries are mainly financed by general tax revenues and/or health-related public insurance programs and provided as a kind of social or health service. Thus it is often hard to find exogenous sources of variation in LTC services that enable researchers to identify fiscal, economic and social consequences of public LTC programs.

Second, even if we find a distinct introduction or expansion of a public LTCI program, it is difficult to estimate its causal impact due to the universality of current LTCI schemes in several countries. In short, because there are no solid "control" groups within the same country due to the universality of LTCI programs, we cannot compare socioeconomic outcomes of those who are covered by LTCI to their estimated counterfactual

[1] See, among others, Wolf and Soldo (1994), Ettner (1995; 1996), Carmichael and Charles (1998; 2003a;b), Heitmueller (2007), Heitmueller and Inglis (2007), Bolin et al. (2008b), Johnson and Lo Sasso (2006), Leigh (2010), Lilly et al. (2010), Michaud et al. (2010), Moscarola (2010), Casado-Marín et al. (2011), Ciani (2012), Kotsadam (2011), Meng (2012; 2013), Van Houtven et al. (2013), Crespo and Mira (2014), Nguyen and Connelly (2014), Skira (2015), Yamada and Shimizutani (2015), Oshio and Usui (2017; 2018), and Schmitz and Westphal (2017). Lilly et al. (2007) and Bauer and Sousa-Poza (2015) review studies on the impact of informal caregiving on caregivers' labor supply and related outcomes. The estimated effects of informal caregiving on the (female) labor supply in these studies are different and heterogeneous, but both Lilly et al. (2007) and Bauer and Sousa-Poza (2015) conclude that in general estimated negative impacts tend to be small or modest.

[2] Recent studies include Van Houtven and Norton (2004), Charles and Sevak (2005), Hanaoka and Norton (2008), Bolin et al. (2008a), Bonsang (2009), Barczyk and Kredler (2017) and Barczyk and Kredler (2018).

outcomes. This universal feature of existing public LTCI is a major obstacle to the plausible identification of the impact of an LTCI introduction.[3]

To overcome the difficulty of finding a reliable control group within the same country, we estimate the nationwide aggregate impact of a large-scale LTCI introduction in Japan on public finance and female labor force participation, utilizing within-country variations in country-level panel data. Our empirical strategy relies on the synthetic control (SC) method developed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) for plausible statistical causal inference in a case study. By "case study"" we mean that the number of the "treated" cases or units is only one, which in this paper is Japan.

Our findings suggest that LTCI introduction substantially increased in-kind benefits in Japan but did not crowd out public health expenditure. We also do not find any positive LTCI impact on labor force participation for middle-aged women. These findings imply that LTCI introduction in Japan was not a sufficient booster capable of altering Japanese female-dependent informal caregiving and low female labor market participation, which are often identified as characteristics of Japanese familialism (OECD 2012; 2017).

Our contributions are three-fold. First, this is to our knowledge the first study that investigates the nationwide general-equilibrium impacts of a large-scale LTCI introduction. Most previous studies of LTCI effects on labor supply, which we will discuss in Section 2.3, use individual-level data to identify partial-equilibrium effects, explicitly or implicitly investigating changes in the labor supply of informal caregivers before and after LTCI introduction.

These micro-level partial effects are informative and policy-relevant, but they do not provide information on how a nationwide universal LTCI introduction has (or has not) changed the country in question's aggregate fiscal and labor-market conditions.[4] Our result of no LTCI effect on female labor force participation is different from some micro-level empirical evidence and this suggests that we need to reconsider several possible pathways from LTCI to female labor supply.

Second, while some recent influential historical or cross-country studies on the determinants of female labor force participation do not focus on the roles of informal and formal LTC (Goldin 2006; 2014; Olivetti and Petrongolo 2016; 2017), some cross-country or within-country studies find a negative relationship between the level of "family ties" or "home production" and (female) labor force participation.(Alesina and Giuliano 2010; Ngai and Pissarides 2011) It is therefore interesting to shed light on the nationwide impact of a universal LTCI program on female labor force participation because a large-scale LTCI program could alter the balance between home production and publicly subsidized LTC services.

[3] We discuss how previous studies try to avoid this problem in Section 2.

[4] See, among others, Heckman et al. (1998), Blundell et al. (2004) and Finkelstein (2007) for the distinction between a partial-equilibrium effect and a general-equilibrium effect in social program evaluations.

In particular, whereas many micro-level studies about the effects of informal caregiving on female labor supply did not find strong negative effects (see footnote 1), Crespo and Mira (2014) found a clear North-South gradient (in Europe) in the positive effect of parental ill health on the probability of informal caregiving by daughters and also observed weaker evidence of a North-South gradient in the negative effect of informal caregiving on female labor force participation.

Although the Japanese case was not studied in Crespo and Mira (2014), the literature of comparative welfare states often categorizes Japan among "familialistic" welfare states, other examples of which include southern and continental European countries, where female family members play primary roles in the provision of child and elderly care (Esping-Andersen 1997; 1999). It was thus expected that Japan would be on the "south" side and the expansion of formal LTC services by LTCI would reduce the burden of female caregivers and boost female labor force participation. The fact that we did not find such an effect at an aggregate level suggests that we need to reexamine the determinants of the labor supply of middle-aged women.

The third contribution of our research is to the literature of country-level synthetic control analysis. While it may be harder to construct a valid synthetic control unit using country-level data because heterogeneity among countries seems large, there is now an increasing number of studies that investigate the aggregate impacts of nation-level reforms on relevant outcomes using country-level panel data and the synthetic control method.[5] To cope with the inherent vulnerability of constructing a synthetic control unit using country-level data, we propose and provide extended sensitivity and placebo analyses.

The rest of the paper consists of the following sections. In Section 2, we discuss the institutional backgrounds of LTCI introduction in international and Japanese contexts. Section 3 explains our empirical strategy with a synthetic control method. In Section 4, we describe our data sources and data arrangements and then show descriptive statistics. Section 5 provides the results of synthetic control estimation and conventional placebo tests. In Section 6, we propose and implement extended placebo analyses. Section 7 discusses our results and concludes our paper.

[5]In the field of health and public policy, Rieger et al. (2017), among others, studied the impact of the introduction of Universal Health Coverage (UHC) on several health spending indicators and infant and child mortality in Thailand in 2001. Barlow (2018) and Olper et al. (2018) examined the impact of trade liberalization on child mortality. Arnold and Stadelmann-Steffen (2017) estimated the effects on social spending of the introduction of federalism in Belgium in 1993. Ryan et al. (2016) investigated the effects on mortality of the introduction of the UK's Quality and Outcomes Framework in 2004. Podestà (2017) studied the impact of family reforms in France and Belgium on female labor force participation. Restrepo and Rieger (2016) analyzed the impact of Denmark's food policy on the mortality rate.

2 Background

2.1 LTCI in the international context

In 2014, the aging rate among OECD nations reached the range of between 6.9% (Mexico) and 27.0 % (Japan) and the averages of the aging rates in OECD members and EU members were 16.8% and 19.8% respectively. All of these numbers are unprecedentedly high (World Development Indicators, 2017). Faced with a situation in which their societies are aging, OECD nations have introduced and developed long-term care (LTC) systems that are based on their own institutional and historical backgrounds.

Table 1 summarizes the characteristics of LTC systems for the elderly among OECD countries in terms of coverage, benefits, and sources of funding based on Colombo et al. (2011).[6] As can be seen in this table, LTC systems are quite diverse among OECD nations, but there are some clusters. First, Nordic countries, which are often considered as leading welfare states, finance LTC costs through tax revenues. In addition, these countries provide LTC services to people with a disability without specific age-related criteria. The U.K., Spain, and the Czech Republic are also categorized in this cluster. Second, many continental European countries such as France, Italy, and Austria adopt more mixed financing systems, but also provide LTC services without strict age-related criteria. Third, public LTC *insurance* (LTCI) has been adopted by only a few continental European and Asian countries such as Germany, Luxembourg, the Netherlands, Japan, and South Korea, where public health insurance systems had already been adopted before the introduction of LTCI.[7]

Table 1: LTC systems in OECD countries

	Coverage and benefits			
	People with a disability		Aged people with a disability / People with an age-related disability	
Sources of funds	In kind	Cash and in kind	In kind	Cash and in kind
Tax revenues	Canada	Czech Republic, Denmark, Finland, Ireland, Norway, Spain, Sweden, UK	Greece	Slovak Republic
LTC insurance (Premiums and taxes)		Germany, Luxemboug, Netherlands	**Japan**	Korea
Mixed	Hungary, Portugal	Austria, Belgium, France, Italy Poland, Slovenia, Siwitzerland	Australia	Mexico, US

Source: the author's tabulation based on Colombo et al. (2011), Table 7.1.

[6]We do not consider LTC systems that target only non-elderly people with a disability.

[7]From a more dynamic, rather than comparative, perspective, Colombo et al. (2011) states that "over time, coverage systems are evolving towards universal systems or benefits and more user-choice models, with, in many cases, increased targeting of care benefits to those with the highest care needs."

Japan has had LTCI since 2000. One important feature of Japanese LTCI is that LTCI introduction has caused sharp, not incremental, increases in LTC financing and spending. This provides us with a good opportunity to identify the impact of a large-scale LTCI introduction.

2.2 LTCI in Japan

Before LTCI was introduced in 2000, public LTC services in Japan were mainly means-tested programs for the low-income elderly. Under the means-tested programs, the elderly people in need of LTC but ineligible for public LTC benefits were often admitted to hospitals and stayed there for a long time even after necessary medical treatment had concluded (Campbell and Ikegami 2000; 2003).

This is called "social hospitalization" of the elderly, which was (and still is) considered a notorious social phenomenon in Japan's aging society. This problem was exacerbated by the introduction of a new health care scheme for the elderly in 1983, which had a relatively generous payment system for elderly hospital admission. In order to minimize such "social hospitalization" and to cope with both increasing medical costs for the elderly and the expanding need for LTC services caused by a rapidly aging population, in the 1990s the Japanese government implemented several reforms that were financed by national and local taxes. Due to several limitations of the tax-financing LTC system, the Long Term Care Insurance Law was enacted in 1998 and enforced in April of 2000.

In what follows we explain the institutional setting of LTCI in Japan based on Campbell et al. (2010) while comparing it with LTCI in Germany[8]. The comparison with Germany is meant to clarify the characteristics of Japanese LTCI compared with German LTCI, which was the first LTCI in an OECD country. Based on Table A.1 in Appendix A, which simply replicates Exhibit 3 of Campbell et al. (2010), we selectively discuss the financing, population coverage, eligibility, and benefits of LTCI in Japan and Germany. All the information below is based on institutional settings in 2008. For more details about the institutional settings of LTCI in Japan and Germany, see Campbell (2002), Campbell et al. (2010) and Rhee et al. (2015).

To begin with, LTCI in Japan is managed as a uniform and independent social insurance system, but is financed by several fiscal resources such as insurance premiums and taxes, whereas in Germany LTCI is financed only by insurance premiums. See Table A.1 for more detailed information about LTCI financing in Japan.

Second, in both Japan and Germany LTCI is a universal program that does not require means testing for eligibility for LTC services. Population coverage, however, does differ between the two programs. In Germany, people of all ages are eligible for LTCI

[8] See also Campbell and Ikegami (2000), Campbell and Ikegami (2003) and Campbell et al. (2010) for the historical, institutional, and political backgrounds of LTCI introduction in Japan.

benefits. In Japan, all people aged 65 and over are covered by Japan's LTCI, but otherwise only people aged 40-64 who have age-related diseases are eligible for LTCI benefits. In Japan, LTC benefits for younger people with a disability are mostly provided by local governments and financed by tax revenues.

Third, care-need assessments and eligibility criteria also differ between Japan and Germany. According to Table A.1, percentages of the eligible elderly and the benefit-receiving elderly were higher in Japan than in Germany in 2008.

Finally, the types of LTCI benefits also differ considerably between Japan and Germany. One of the most noticeable differences is that Germany's LTCI has cash benefits but Japan's LTCI does not, providing only in-kind benefits. According to Figure A.2 in Appendix A, which is based on Exhibit 2 in Campbell et al. (2010), spending on cash allowances for informal caregiving is the second largest LTC expenditure item in Germany. On the other hand, Japan has more generous in-kind benefits. In Figure A.2, per capita LTCI expenditures for all in-kind benefits in Japan outweigh those in Germany.

The Japanese government did not include cash benefits for informal caregivers in LTCI partly because of opposition from women's caregiver groups who argued that cash benefits could have strengthened female gender roles in informal caregiving and prevented women from joining or staying in the labor market. Although Hieda (2012) indicates that the Ministry of Health and Welfare excluded the option of cash benefits in the early stages of the policy-making process due to fiscal reasons, it is interesting that there existed a social movement that explicitly opposed cash benefits from a gender perspective.

2.3 Previous literature

How has LTCI introduction in Japan affected fiscal and socioeconomic outcomes? First, one obvious assumption is that public expenditure for LTC will have increased. Second, what is relatively unclear is whether or not crowding-out effects on other public expenditures exist. For example, public health care expenditure is expected to decrease if newly introduced LTC benefits are substitute goods for some other elements of Japan's health care services. In fact, as discussed before, one policy objective of Japan's LTCI introduction was to reduce unnecessary "social hospitalization".

The effects of LTCI introduction on non-fiscal socioeconomic outcomes should be very diverse. Using individual-level survey data, previous literature in epidemiology, economics, and other social sciences has tried to identify the impacts of LTC services on various outcomes such as LTC service utilization, the well-being indices of care receivers and family caregivers, and the labor supply of family caregivers.

In this paper we focus on LTCI's macro-level impact on female labor participation. The advantage of our using country-level data is that we can examine the nation-level general-equilibrium impact of LTCI, which is rarely investigated in the literature.

Previous individual-level studies of LTCI effects on female labor supply present mixed results. Using structural models and German data, Geyer and Korfhage (2015) find that LTCI benefits in kind in Germany have small positive effects on labor supply, LTCI cash benefits have larger negative effects, and the average effect is significantly negative. Exploiting a difference-in-differences strategy, Geyer and Korfhage (2018) examine the labor supply effects of LTCI introduction in Germany in 1995 for co-residential carers and find some negative effect for men and no statistically significant effect for women. Shimizutani et al. (2008) find no effect of LTCI introduction in Japan on female labor market participation in 2001, just one year after LTCI introduction, but a large positive effect in 2002. Tamiya et al. (2011) finds some positive but heterogeneous LTCI effects on family carer labor supply, with high-income households receiving a higher impact. Sugawara and Nakamura (2014) find positive effects of LTCI introduction and diffusion in Japan on female labor supply. Fu et al. (2017) also find positive LTCI effects on the labor supply of male and female family caregivers. On the other hand, Fukahori et al. (2015) argue that LTCI introduction did not mitigate the negative impact of family care needs on household labor supply.[9]

Overall, previous studies imply that LTCI with in-kind benefits may have some positive effect on female labor supply, whereas LTCI with cash benefits seems to have a negative effect, although evidence is still insufficient to draw a strong conclusion. If these implications based on individual-level studies can be straightforwardly applied to a macro-level analysis, we expect Japan, where only in-kind benefits are available, to have experienced a positive LTCI impact on female labor supply.

The findings of the above micro-level studies are very important, but there are some limitations. Several previous studies argue that they utilize a difference-in-differences (DID) method as their identification strategies (Geyer and Korfhage (2018) for Germany, Shimizutani et al. (2008), Tamiya et al. (2011), Fukahori et al. (2015) and Fu et al. (2017) for Japan). Treatment and control groups in these studies, however, are not defined based on an exogenous group-level exposure to LTCI introduction as a standard DID framework implies. This is because in Germany and Japan LTCI programs were uniformly introduced nationwide and their coverage is universal (for all generations in Germany and for the elderly in Japan). Hence it is impossible to define the control group as *informal caregivers* who are not affected by LTCI introduction.

Most of the previous studies therefore define "being an informal caregiver" or "living with frail elderly" in the post-LTCI introduction period as a "treated" status. This implies that they essentially try to identify how LTCI introduction reduces the negative effect of "being a informal caregiver" on the labor supply by comparing the coefficients

[9]As a relevant study, Yamada and Shimizutani (2015) find that being eligible for LTCI services partially mitigates the negative impact of caregiving on the labor supply in Japan. Kondo (2017) investigate the effect of long-term care facilities on the labor supply of middle-aged people in Japan and find no evidence of such an effect.

of the "caregiver or not" dummy variable before and after LTCI introduction. This empirical strategy is useful particularly because a direct comparison between LTCI-affected and non-LTCI-affected caregivers is impossible under the introduction of universal and uniform LTCI in Japan and Germany.

The introduction of a universal LTCI scheme, however, should also affect female labor supply by influencing, among other factors, the decision-making behind "being a caregiver or not" (or "living with frail elderly or not") itself. This may result in possible endogeneity bias and the violation of SUTVA (Stable Unit Treatment Value Assumption) in such a research design.

Individual-level endogenous determination of informal caregiving is well recognized in the literature, but large-scale LTCI introduction can also affect female labor force participation through the creation of employment opportunities for middle-aged women.[10] This indirect effect of LTCI introduction may lead to different results and implications between studies in partial-equilibrium and general-equilibrium (or micro-level and macro-level) frameworks.

One alternative way to identify the causal effect of an LTCI introduction that takes into account these problems is to exploit some regional variation in the intensity of the LTCI introduction. For example, Løken et al. (2016) find a negative impact of formal LTC expansion on insured work absences for the adult daughters of single elderly parents in Norway. Their identification strategy is to exploit the differential increase in the availability of federal funds in municipalities caused by a national LTC reform for the elderly. The application of such an identification strategy requires finding some regional or other variations in the intensity of LTCI introduction. It is however hard to find such an exogenous variation in the introduction of a universal and uniform LTCI, which may explain why previous studies in Japan and Germany utilize the different identification strategies described above.

We therefore shift our focus from a micro-level or municipality-level variation to a country-level variation to examine the aggregate impact of LTCI introduction. Although there are some drawbacks in exploiting cross-country variation for the causal inference in general, our empirical analysis will provide useful implications concerning LTCI impacts.

3 Empirical strategy

3.1 A case study using the synthetic control method

Because our study focuses on the specific nationwide event of LTCI introduction in Japan using country panel data, we have only one "treated" unit in our sample for analysis. The

[10] For example, some empirical welfare-state studies such as Mandel and Semyonov (2006) emphasize the role of the welfare state as a provider of employment opportunities for women.

synthetic control (SC) method proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) is a suitable method to investigate the impact of such a single but noticeable event. We will now briefly explain how the SC method achieves the identification of aggregated LTCI effects.

First, let's define the aggregate effect of the LTCI introduction as α_{it} on some outcome variable Y_{it}, where i and t indicate a country and a year respectively. This implies we assume that the effect of the LTCI introduction varies across countries and years. Next, we consider the situation in which an LTCI program is introduced in country $i = J$ (i.e. Japan) in year T_0 and assume that the LTCI introduction is fully implemented and irreversible. In this case, we can define the treatment effect α_{Jt} as follows:

$$\alpha_{Jt} = Y_{Jt}(1) - Y_{Jt}(0), \quad \text{for } t > T_0 \tag{1}$$

where $Y_{Jt}(D_J)$ is a potential outcome with the intervention status D_J, where $D_J = 1$ indicates the LTCI introduction and $D_J = 0$ represents no LTCI introduction. Thus $Y_{Jt}(1)$ is identical to an observed outcome Y_{Jt}^{obs} and $Y_{Jt}(0)$ is a "counterfactual" outcome that would be realized if country J did not introduce LTCI in $t \geq T_0 + 1$. In order to estimate α_{Jt}, we need to estimate $Y_{Jt}(0)$ in $t \geq T_0 + 1$.

Abadie and Gardeazabal (2003) and Abadie et al. (2010) propose a novel method to estimate $Y_{Jt}(0)$ by utilizing the weighted average of outcome variables of control units i $(i = 1, 2..., N)$, that is $\sum_{k \neq J} w_k^* Y_{kt}^{obs}$. An optimal time-invariant weight w_k^* for each control unit k is determined so that the vector of optimal weights $W^* = (w_1^*, w_2^*, ..., w_k^*)'$ minimizes the difference between the pre-intervention outcomes and characteristics (called predictors) of the treated unit and the weighted average of predictors of the control units, given that $0 \leq w_k^* \leq 1$ and $\sum_{k \neq J} w_k^* = 1$. A single fictional control unit constructed by the optimal weights W^* is called a synthetic control.

Thus, a synthetic control has pre-intervention outcomes and characteristics which are set as similarly as possible to those of the treated unit in terms of observed predictors, but it does not receive a treatment in the post-intervention period. Therefore the outcome of the synthetic control in the post-intervention period is meant to represent the counterfactual status of the treated unit $Y_{It}(0)$.

Given that the synthetic control can provide unbiased estimates of the counterfactual status of the treated unit $Y_{It}(0)$, α_{It} is estimated as follows:

$$\hat{\alpha}_{Jt} = Y_{Jt}^{obs} - \sum_{k \neq J} w_k^* Y_{kt}^{obs}. \tag{2}$$

Building on some parametric assumptions but allowing for time-varying unobserved confounders, Abadie et al. (2010) prove that the above SC estimator is unbiased if the treated unit and the synthetic control are well matched in observed predictors and outcome variables in long pre-intervention periods.

In a subsequent study, Abadie et al. (2015) recommend that the SC method should be applied in cases where a sizable number of pre-intervention periods are available in order to construct a credible synthetic control. We examine the effects of Japanese LTCI introduction since 2000 on fiscal outcomes and female labor force participation. Our pre-intervention periods are in most cases about 20 years (1980-1999).

3.2 Informal test of the null hypothesis

One weakness of the SC method is that it does not provide a formal statistical test for the null hypothesis. As a complement to formal statistical hypothesis testing, Abadie et al. (2010) provide an alternative, informal placebo test akin to a permutation or randomization test in which a researcher calculates and collects "placebo" SC estimates by assigning the "label" of the intervention status to each control unit and then compares a true SC estimate to these placebo values. Most of the previous studies using the SC method show the results of this kind of placebo test, and we also present the results of this conventional test. In Section 6, we further explore the placebo analysis in the SC method and provide extended placebo trials that are still informal but more rigorous and we hope more informative.

3.3 Selection of donor pool countries

One important issue in SC analysis is how to select the candidates for control countries, which are called "donor pool" countries. Due to data availability, we limit donor pool countries to OECD nations. This data restriction is justifiable because it is preferable to have relatively homogeneous control units in a donor pool that are reasonably comparable to the treated unit in terms of socio-economic characteristics (Abadie et al. 2010; 2015).

In addition, we exclude Germany, the Netherlands, Luxembourg and South Korea from the donor pool because these countries adopted LTCI during the sample period. This means that we do not allow these countries to be included in the synthetic Japan.

3.4 Rationales for case studies

Some may wonder whether it is plausible to use country-level macro data and the SC method to investigate the impact of LTCI introduction on fiscal and labor outcomes. One obvious drawback of the combination of using country-level data and the SC method is the difficulty of constructing a plausible counterfactual situation based on other countries using this approach. We recognize this limitation but argue the presentation of our empirical results is still worthwhile for the following reasons.

First, by exploiting cross-country variation we can examine the general-equilibrium nationwide impact of LTCI introduction that incorporates both direct and indirect responses of demand and supply sides in the national long-term care markets. This is something that individual-level and regional-level studies, in particular the former, cannot obtain.

In addition, the research designs of the previous individual-level studies assume that "untreated" individuals (i.e. non-caregivers) are not affected by LTCI introduction. As already discussed, however, the choice of being non-caregivers may be affected by the introduction of LTCI, and non-caregivers may also be affected by the expansion of employment opportunities caused by LTCI introduction. The possibility of these indirect LTCI effects implies the risk of a violation of SUTVA (Stable Unit Treatment Value Assumption) in the identification strategies of the previous studies. Our cross-country comparisons are instead made on more straightforward counterfactual outcomes based on "untreated" countries that are not affected by Japanese LTCI introduction.

Second, SC analysis is perhaps one of the most plausible and transparent methods for a comparative case study. As we discuss further in Section 6, the SC method is vulnerable to an imbalance of pre-intervention predictors and idiosyncratic shocks in both treated and control units. Nonetheless, because SC analysis provides us explicit information about how the synthetic control is constructed using weights W^*, we can use complementary qualitative assessments of the treated country and high-weight control countries to examine obtained SC estimates and their implications.

4 Data

For our empirical analysis, we construct annual panel data for 22 OECD countries from 1980 to 2013 by combining various data sources. Table B.1 in Appendix B presents a complete list of the definitions and sources of our data set.

To begin with, our main fiscal outcome to be investigated is the variable of in-kind benefits for the elderly because Japanese LTCI provides only in-kind benefits and covers only the elderly. In order to investigate the crowding-out effects of LTCI on other related public expenditures, we then also collect data for public health expenditures that in principle do not include long-term care expenditure. Finally, we use six variables describing female labor supply: female labor force participation (LFP) rates for middle-age cohorts 30-34, 35-39, 40-44, 45-49, 50-54 and 55-59. Unfortunately, we cannot analyze the counterpart male LFP rates with the SC method because Japanese male LFP rates are among the highest in the OECD countries and a valid "synthetic Japan" cannot be constructed based on other OECD countries.

Our main predictors are pre-intervention outcomes and demographic variables. When it comes to pre-intervention outcomes, all of these are used as separate predictors based on the theoretical and empirical findings of Ferman et al. (2018). Demographic variables consist of population under 15 as % of the total population (child population), the growth rate of the child population, population aged 65 and over as % of the total population (elderly population), and the growth rate of the elderly population. These data are obtained from OECD Employment and Labour Force Statistics. Other demographic variables are employment in agriculture (% of civilian employment), employment in industry, and employment in services. These data come from the "Comparative Welfare State Dataset" (Brady et al. 2014).

We also include additional predictors that are meant to capture the impact of economic development on the outcomes of interest: per capita GDP and GDP growth. We use expenditure-side real GDP, which is taken from the "Pen World Table8.1" (Feenstra et al. 2015). Per capita GDP is calculated as expenditure-side real GDP divided by population.

Table 2 presents descriptive statistics of our OECD panel data. Original data consist of unbalanced panel data for 18 OECD countries.[11] between 1977 and 2013, although the data availability significantly differs among years and countries. In a subsequent analysis, we use data after 1980. In order to implement SC estimation with annual data we impute missing values by linear interpolation, but we do not extrapolate any values. Thus we sometimes drop years or countries due to data limitations depending on the outcome variable.

5 Results

5.1 Impacts on in-kind benefits for the elderly

Figure 1 provides the results of SC estimation for in-kind benefits for the elderly. Thick solid lines are realized in-kind benefits as % of GDP and the other three lines are the counterpart values of three synthetic controls.

Synthetic control 1 in the graph is constructed from the original donor pool and therefore its values are regarded as baseline counterfactual outcomes in the post-intervention period, that is, they represent the levels of in-kind benefits for the elderly if LTCI had not been introduced in Japan.[12]

[11]Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Italy, Japan, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. Greece is excluded from the sample because of its unusual budgetary situation in the late 2000s. We also exclude Germany, the Netherlands and Luxemburg because like Japan they too have adopted LTCI schemes. We also drop Eastern European countries and South Korea due to data limitations.

[12]As discussed in Section 3, SC estimates are the gaps between the outcomes of a treated unit and a synthetic control. If a synthetic control is validly constructed based on pre-intervention outcomes and predictors, SC estimates are expected to be around zero in the pre-intervention period and can be interpreted as causal effects in the post-intervention period.

Table 2: Descriptive statistics

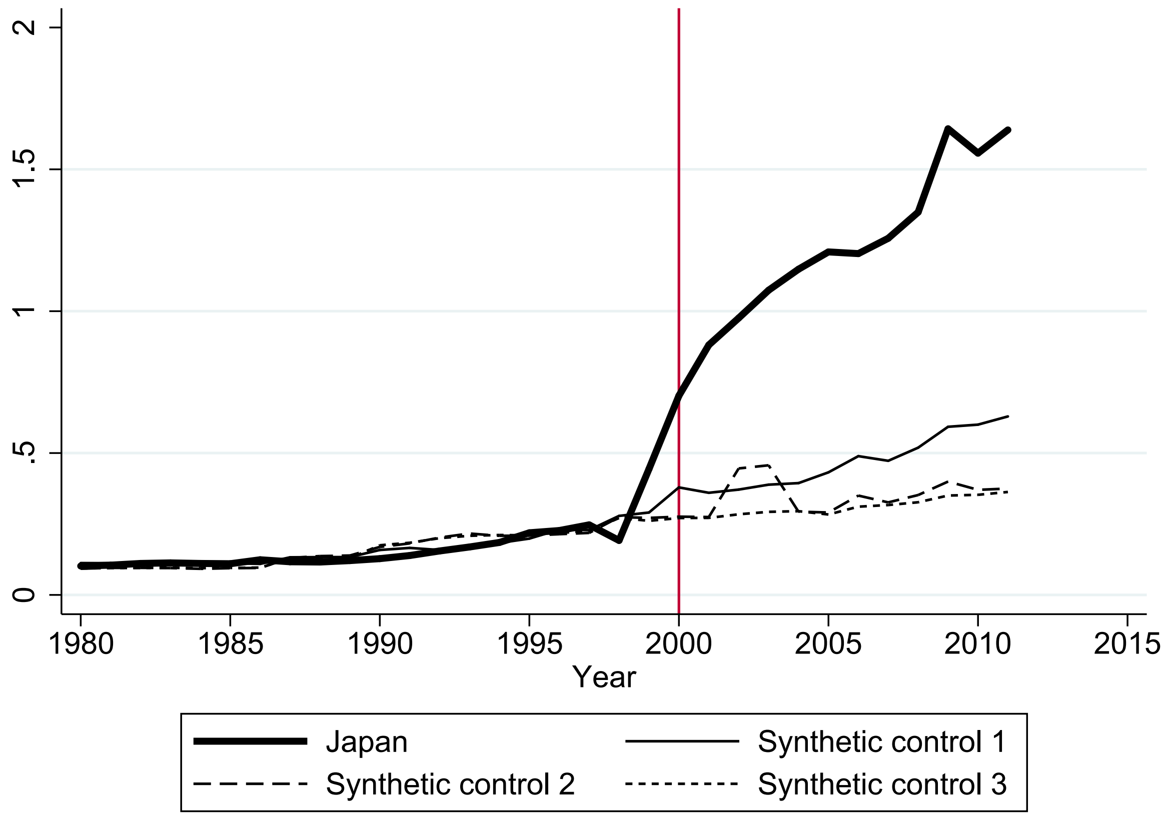
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Outcomes					
Public expenditure on benefits in kind for the elderly (% of GDP)	515	0.613	0.768	0.000	2.912
Total public expenditure on health care (% of GDP)	548	5.523	1.286	0.000	8.733
Total public social expenditure (% of GDP)	576	21.102	5.449	9.643	35.517
Female labor force participation rate (age 30-34)	531	73.928	10.692	29.536	91.639
Female labor force participation rate (age 35-39)	522	74.894	10.874	28.151	92.361
Female labor force participation rate (age 40-44)	522	75.940	11.996	27.298	93.538
Female labor force participation rate (age 45-49)	522	73.866	13.883	26.467	92.734
Female labor force participation rate (age 50-54)	522	67.057	15.879	24.409	88.542
Female labor force participation rate (age 55-59)	531	53.063	17.398	14.414	83.409
Predictors					
Per capita real GDP	544	26.547	7.765	8.244	53.896
Child population (%)	578	18.788	2.654	12.875	27.195
Elderly population (%)	578	14.792	2.519	9.100	25.058
Employment in agriculture (%)	544	5.989	4.164	1.032	27.259
Employment in industry (%)	544	27.567	5.464	0.000	40.261
Employment in services (%)	544	65.721	9.867	0.000	81.236
Growth rate of per capita real GDP	544	2.109	3.269	-10.263	14.856
Growth rate of population	544	0.589	0.486	-0.883	3.928
Growth rate of child population	578	-0.942	1.001	-4.153	1.404
Growth rate of elderly population	578	1.073	1.125	-3.027	4.909

Notes: Original data is unbalanced panel data for 18 OECD countries between 1980 and 2013. In order to implement synthetic control analysis with annual data we impute missing values by linear interpolation, but we do not extrapolate any values. OECD countries include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Italy, Japan, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. Greece is excluded from the sample because of its unusual budgetary situation in the late 2000s. We also exclude Germany, the Netherlands and Luxemburg because like Japan they too have adopted LTCI schemes.

Synthetic control 2 is constructed using SC estimation in which the country that receives the highest weight in the first SC estimation is excluded from the donor pool. Synthetic control 3 is constructed from a donor pool that additionally excludes the country that receives the highest weight in the second SC estimation. These robustness checks are particularly important in our cross-country comparison where there is a risk that some specific countries receive higher weights and idiosyncratic shocks in these countries may undermine the validity of the SC estimation. See also Abadie et al. (2015) for further discussion of this type of sensitivity checks. In the online appendices I, we provide the weights and pre-determined covariate values used for constructing synthetic controls 1,2, and 3.

The results of SC estimation in Figure 1 indicate sharp increases in in-kind benefits for the elderly in Japan just after the introduction of LTCI. The gaps between the actual benefit level and those of the synthetic controls persist and increase during the sample period and the size of the gaps reaches around one % point of GDP in 2011.

Figure 1: SC estimation for in-kind benefits for the elderly (% of GDP)

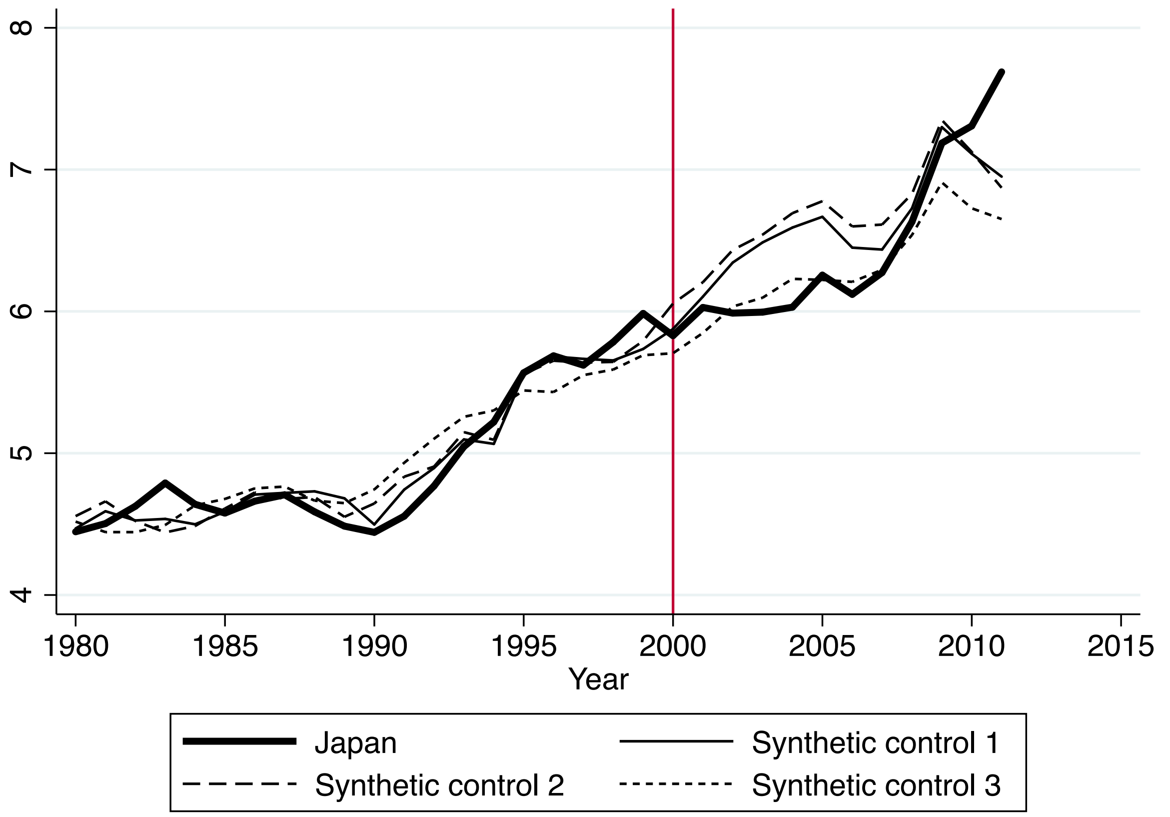


Notes: Synthetic control 1 is constructed from the original donor pool, synthetic control 2 is constructed from the donor pool that excludes the country that receives the highest weights in the first SC estimation, and synthetic control 3 is constructed from the donor pool that also excludes the country that receives the highest weights in the second SC estimation. Canada is excluded from the original donor pool due to lack of data. For SC estimation we use the `synth` command in Stata with the `nested` and `allopt` options. See online Appendix I for detailed estimation results.

5.2 Crowding out health expenditures?

We then examine whether LTCI introduction crowds out closely related public expenditure, that is, other public health expenditures. Figure 2 provides the results of SC estimation for public health expenditure as % of GDP. When we compare actual outcomes with those of the synthetic controls 1 and 2, the gaps between outcomes in Japan and synthetic Japan are negative in the early 2000s, indicating that LTCI introduction might have led to the suppression of public health expenditure in this period. This suppression, however, cannot be robustly observed when the synthetic control 3 is used as a counterfactual unit. Overall, there is no clear evidence that LTCI introduction has caused a public-expenditure shift from health care to long-term care and the persistent suppression of public health expenditure.

Figure 2: SC estimation of public health expenditure (% of GDP)



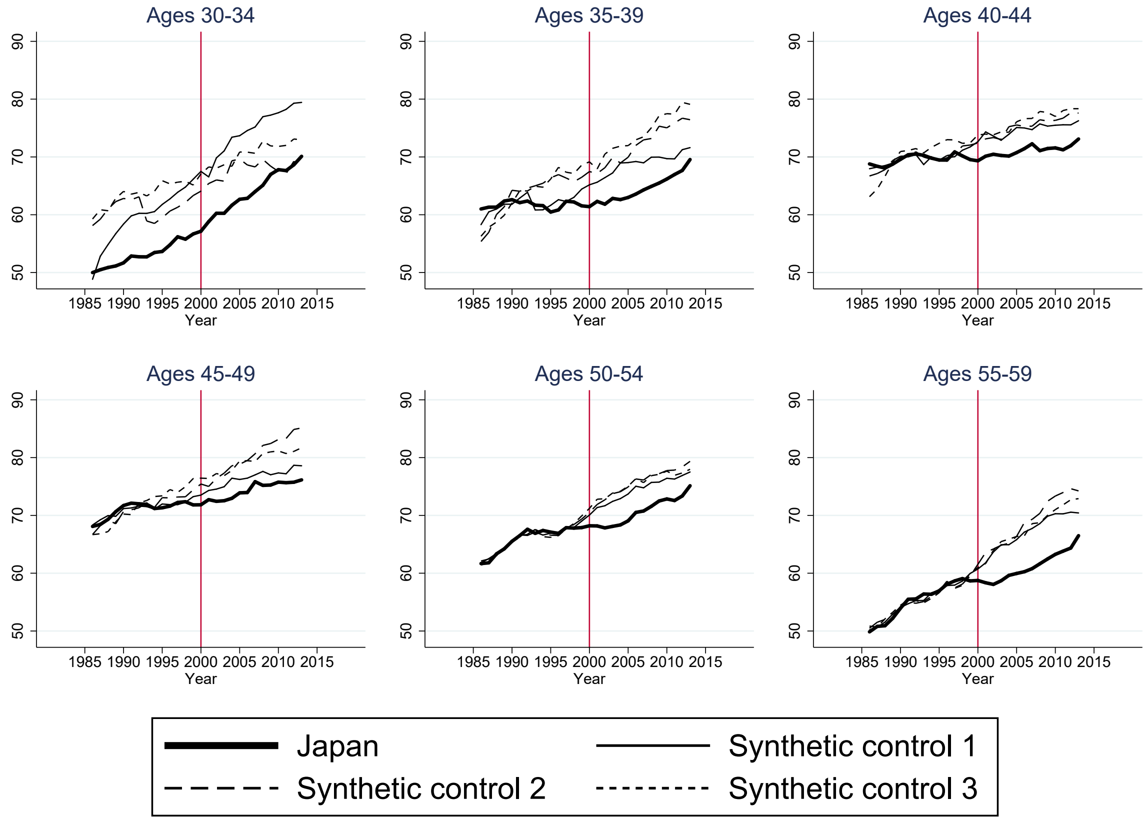
Notes: Synthetic control 1 is constructed from the original donor pool, synthetic control 2 is constructed from a donor pool that excludes the country that receives the highest weights in the first SC estimation, and synthetic control 3 is constructed from a donor pool that also excludes the country that receives the highest weights in the second SC estimation. Norway is excluded from the original donor pool due to a lack of data. For SC estimation, we use the `synth` command in Stata with the `nested` and `allopt` options. See online appendices I for detailed estimation results.

5.3 Impact on female labor force participation

Moving on from fiscal outcomes, Figure 3 provides our SC estimation results for female labor force participation (LFP) rates by age cohort. Despite large fiscal expansion for LTCI , there is no sign of positive LTCI effects on the LFP rates in any of the cohorts. In fact, female LFP rates appear to even have been suppressed after LTCI introduction compared with those of all of the synthetic controls. Note also that Japan's LFP rates for the youngest cohort of ages 30-34 are lower than all three synthetic controls in the pre-intervention period, implying that it is not possible to construct a valid synthetic control for this cohort from the donor-pool OECD countries.

The overall tendency of Japan's stagnated LFP rates in the post-intervention period suggests that there may exist a Japan-specific trend in the female LFP rates that is not taken into account by the synthetic controls. This implies that SC estimates (i.e. outcome gaps between Japan and a synthetic control) may not properly capture the causal effects of the LTCI introduction.

Figure 3: SC estimation of female LFP rates by age cohort

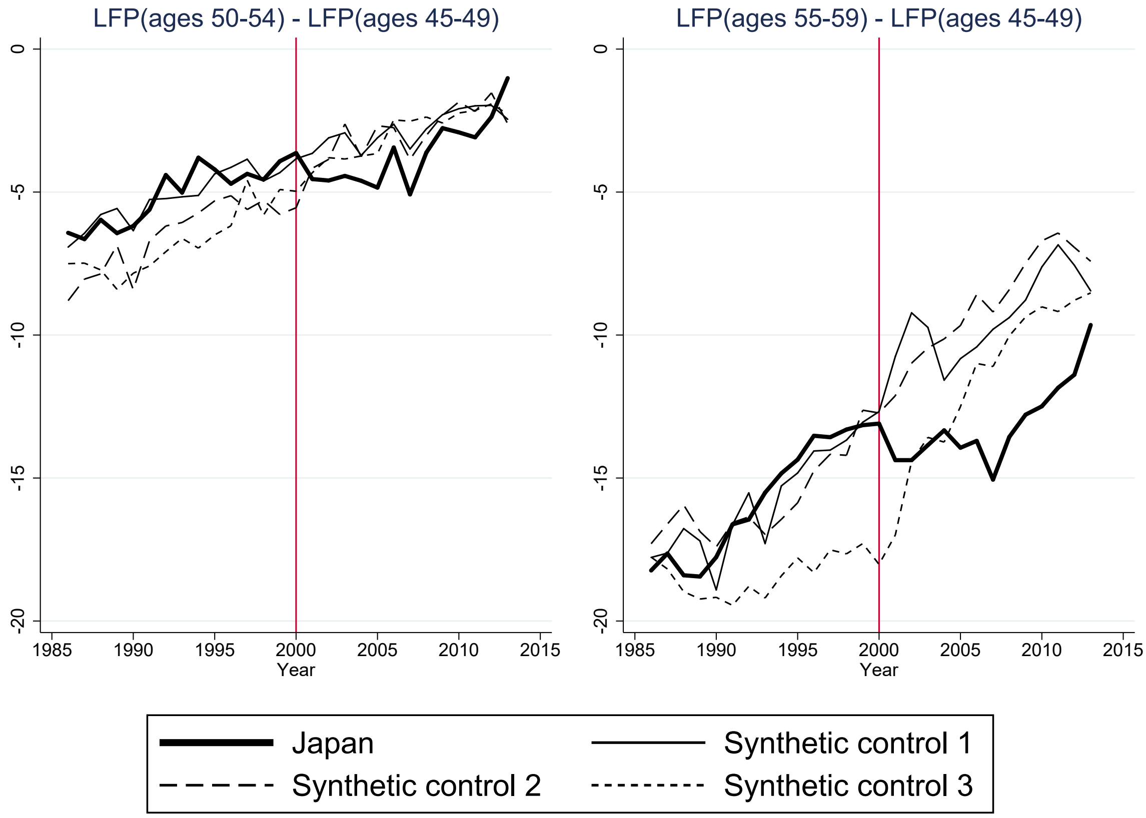


Notes: See the note on Figure 1 for a detailed explanation of the graph. Due to data availability, the first year of our sample is 1986. Austria, Ireland, and Switzerland are excluded from the original donor pool due to lack of data. Except for the age cohort 30-34 (upper left graph), Finland is also excluded due to lack of data. For SC estimation, we use the `synth` command in Stata with the `nested` and `allopt` options, but in cases in which there is an optimization error (due to a poor pre-intervention fit) we implement synth without `nested` and `allopt`. See online appendices I for detailed estimation results.

In order to eliminate this possible Japan-specific trend in female LTF rates, we subtract the female LFP rate of ages 45-49 from the female LFP rate of ages 50-54 and ages 55-59 and then use these differenced variables as outcomes. The idea behind this procedure is that women ages 45-49, whose LFP rate is the highest among the six age cohorts in the post-reform period, are likely to be less affected by LTCI introduction because their parents and parents-in-law tend to still have no need of LTC, whereas women ages 50-54 and 55-59 are likely to be more affected by the LTCI introduction because of higher need for LTC of their parents or parents-in-law. Thus subtracting the female LFP rate of age cohort 45-49 from that of an older cohort may effectively eliminate the Japan-specific trend of female LFP, leaving a change in the older-cohort LFP rate caused by LTCI introduction. This estimation strategy is akin to the triple-difference or difference-in-difference-in-difference strategy, although we use the SC method after differencing the outcomes of “more affected” and “less affected” cohorts in both treated and control (or donor-pool) countries.

Estimation results based on this strategy are shown in Figure 4. The left graph shows the trend of LFP-rate differences between the age cohort 50-54 and the age cohort 45-49 in Japan (bold solid line) and its synthetic controls. The right graph presents the counterpart trends of LFP-rate differences between the age cohort 55-59 and the age cohort 45-49. The results also do not indicate any positive impact of LTCI introduction on female LFP rates for these two age cohorts. In fact, the right-hand graph again shows that the female LFP rates for the age cohort 55-59 seem to be suppressed after 2000 even after eliminating the trend of female LFP for the age cohort 45-49.

Figure 4: SC estimation for the difference in female LFP rates by age cohort



Notes: See the note on Figure 1 for a detailed explanation of the graph. Due to data availability, the first year of our sample is 1986. Austria, Finland, Ireland, New Zealand and Switzerland are excluded from the original donor pool due to lack of data. For SC estimation, we use the `synth` command in Stata with the `nested` and `allopt` options, but in cases in which there is an optimization error (due to a poor pre-intervention fit), we implement `synth` without `nested` and `allopt`. See online Appendix I for detailed estimation results.

5.4 Placebo results

Figure 5 shows estimation results for placebo trials on all of the outcomes except for the female LFP rates of ages 30-34 and 35-39. On the one hand, the first graph in this figure indicates that the SC estimates for in-kind benefits for the elderly seem to be higher than most of the placebo estimates just after 2000, indicating that we can unambiguously conclude there was a fiscal impact of LTCI introduction (around a one % point increase in 2014). On the other hand, we do not find any clear effect on the public health expenditure.

Negative SC estimates for female LFP rates are sometimes clearly larger in size than most placebo estimates. In particular, the female LFP rate for ages 55-59 decreases after 2000 and the magnitude is larger than any placebo estimates. This tendency is mitigated if we subtract the female LFP for ages 45-49 from female LFP for ages 55-59 (the last two graphs). The last graph nonetheless indicates that the female LFP rate for ages 55-59 stagnated after 2000 and the magnitude is larger than most of the placebo estimates, although pre-intervention fits are poor for many placebo trials.

6 Further tests of no effect

6.1 Motivation and Setup

A limitation of the SC method is that this method does not provide formal statistical testing and the conventional placebo tests presented above are beneficial but still crude. In this section, based on the same parametric factor model used in Abadie et al. (2010), we first discuss how a SC estimate and counterpart placebo estimates can deviate from the true parameters of interest. We then provide four test statistics for extended placebo tests based on our discussion, placebo estimation (Abadie et al. 2010) and leave-one-out estimation (Abadie et al. 2015). Finally, we apply our proposed placebo tests to some selective outcomes.[13]

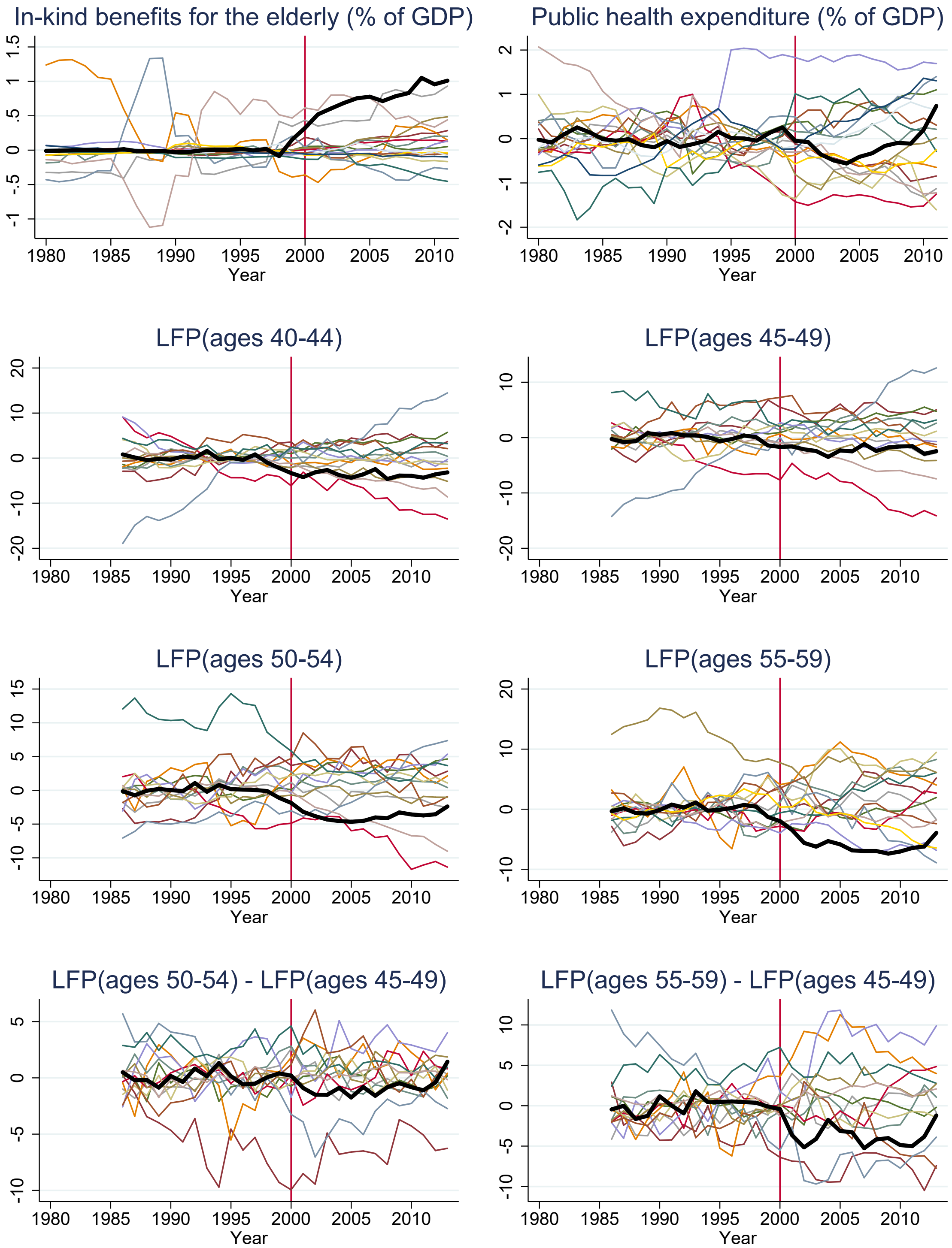
We assume that the data-generating process in the absence of LTCI introduction can be expressed as the following “motivating model” in Abadie et al. (2010):

$$Y_{it}(0) = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \varepsilon_{it}, \tag{3}$$

where $Y_{it}(D_i)$ is the potential outcome defined in Section 3.3, δ_t is an unobserved time effect, $\mathbf{Z}_i$ represents observed factor loadings (or predictors/covariates), $\boldsymbol{\mu}_i$ indicates unobserved factor loadings, $\boldsymbol{\theta}_t$ and $\boldsymbol{\lambda}_t$ are time-varying factors (or coefficients), and ε_{it} is unobserved transitory shocks with zero mean.

[13]Our extended placebo tests may be related to the emerging literature on the rigorous inference of SC methods (Doudchenko and Imbens 2017; Ferman and Pinto 2017; Firpo and Possebom 2018; Chernozhukov et al. 2018), but we emphasize that they are still informal robustness/sensitivity checks.

Figure 5: Placebo results



Notes: We calculate placebo SC estimates by assigning the "label" of the intervention status to each control unit, using all the other control units as a donor pool. Note that the composition of donor pools (control units) are different depending on outcome variables due to data constraints. For baseline SC estimates (bold black line), we use the `synth` command in Stata with the `nested` and `allopt` options. For placebo SC estimates (colored line), we implement `synth` without `nested` and `allopt`, because `nested` and `allopt` options sometimes result in optimization errors in some placebo trials.

Based on the equations (2) and (3), we decompose the SC estimator $\hat{\alpha}_{it}$ as follows:

$$\begin{aligned}
\hat{\alpha}_{it} &= Y_{it}^{obs} - \sum_{k\neq i} w_k^* Y_{kt}^{obs} \\
&= Y_{it}(1) - \sum_{k\neq i} w_k^* Y_{kt}(0) \\
&= \alpha_{it} + Y_{it}(0) - \sum_{k\neq i} w_k^* Y_{kt}(0) \\
&= \alpha_{it} + \boldsymbol{\theta}_t(\mathbf{Z}_i - \sum_{k\neq i} w_k^* \mathbf{Z}_k) + \boldsymbol{\lambda}_t(\boldsymbol{\mu}_i - \sum_{k\neq i} w_k^* \boldsymbol{\mu}_k) + \varepsilon_{it} - \sum_{k\neq i} w_k^* \varepsilon_{kt} \qquad (4)
\end{aligned}$$

The equation (4) shows that an actual SC estimate $\hat{\alpha}_{it}$ can deviate from the true causal effect α_{it} due to the following four components. First, the term $\boldsymbol{\theta}_t(\mathbf{Z}_i - \sum_{k\neq i} w_k^* \mathbf{Z}_k)$ represents a deviation that arises from the failure of observed predictor balancing. Second, $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_i - \sum_{k\neq i} w_k^* \boldsymbol{\mu}_k)$ captures a deviation caused by the failure of unobserved predictor balancing. Third, ε_{it} is by definition the unobserved transitory shocks for a treated unit (i.e. Japan) with zero mean. Fourth, $-\sum_{k\neq i} w_k^* \varepsilon_{kt}$ is the negative of the weighted average of the unobserved transitory shocks for the control units.

Although Abadie et al. (2010) theoretically prove that the bias in the estimator of $\hat{\alpha}_{it}$ goes to zero as the number of re-intervention periods increases under the motivating model (3), there are several reasons why the above four deviations can persist in practice. First, the balance of observed and unobserved predictors, which make $\boldsymbol{\theta}_t(\mathbf{Z}_i - \sum_{k\neq i} w_k^* \mathbf{Z}_k)$ and $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_i - \sum_{k\neq i} w_k^* \boldsymbol{\mu}_k)$ zero under perfect balance, may not be sufficiently achieved due to lack of plausible control units in the donor pool. The problem of insufficient balance should not be overlooked in many comparative case studies where non-zero-weighted control units are often limited to somewhat similar but heterogeneous countries, states, regions or municipalities. Second, the transitory shocks on the treated unit ε_{it} and the control units ε_{kt} can also cause the substantial deviation of $\hat{\alpha}_{it}$ from α_{it} because both ε_{it} and ε_{kt} are not averaged away in SC estimation: ε_{it} remains as it is due to a single treatment unit in a typical case study and the weights w_j^* in $\sum_{k\neq i} w_k^* \varepsilon_{kt}$ are often assigned only to several control units in SC estimation.

Abadie et al. (2010)'s placebo test, which we also used in Section 5.4, can be interpreted as an informal test that investigates the distribution of the sum of the four deviations in equation (4). Assuming that the four deviations are unrelated to the intervention status D_i, the sequential placebo assignment of the "label" of the treatment status to one of the control units is expected to generate the distribution of α_{it} under the sharp null hypothesis of $\alpha_{it} = 0$. This distribution enables us to test whether the causal effect of α_{Jt} for the true treated unit $i = J$ is plausibly non-zero.[14]

[14] An intuitive rationale for this placebo test is that if the absolute values of true SC estimates are larger than most placebo SC estimates, true SC estimates should reflect the causal effects of the intervention.

Abadie et al. (2015) provide another robustness check with "leave-one-out" SC estimation in which they exclude from the donor pool a control unit that receives a positive weight in a baseline SC estimation and then re-implement SC estimation. Because the exclusion of an important control unit may result in an increase in $\boldsymbol{\theta}_t(\mathbf{Z}_i - \sum_{k \neq i} w_k^* \mathbf{Z}_k)$ and $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_i - \sum_{k \neq i} w_k^* \boldsymbol{\mu}_k)$ as well as a change in $\sum_{k \neq i} w_k^* \varepsilon_{kt}$ in equation (4), the robustness of $\hat{\alpha}_{it}$ against this exclusion may support the plausibility of SC estimation. We have already implemented this type of robustness check in our baseline analysis in Section 5.

In this section, in order to overcome the small number of placebo trials due to our small donor-pool size, we combine Abadie et al. (2010)'s placebo test and Abadie et al. (2015)'s robustness check by reassigning the treatment label to a control unit and resampling $N_c - 1$ control units from the original N_c control units in the donor pool. By combining both the placebo assignment and the "leave-one-out" procedure we can obtain a richer distribution of the deviation terms in equation (4) under the sharp null hypothesis. Although this is still another informal placebo test, the number of placebo estimates in each t is $N_c(N_c - 1)$ and this is much larger than the number of estimates in the original placebo test (i.e. N_c).

6.2 Implementation

An actual placebo test is implemented as follows. First, we estimate baseline SC estimates for Japan ($i = J$), $\hat{\alpha}_{Jt}$, and leave-one-out SC estimates for Japan, $\hat{\alpha}_{Jt,-h}$, where h is a control unit that is omitted from the donor pool. Second, we estimate leave-one-out placebo SC estimates for each control unit, which we define $\tilde{\alpha}_{it,-h}$, where $i \neq J$ indicates a control unit to which the placebo treatment is assigned.[15] Third, we compare the SC estimates $\hat{\alpha}_{Jt}$ and $\hat{\alpha}_J t, -h$ to placebo SC estimates $\tilde{\alpha}_{it,-h}$. For example, if most of $\hat{\alpha}_{Jt}$, and $\hat{\alpha}_{Jt,-h}$ are above 95 or 97.5% of $\tilde{\alpha}_{it,-h}$, we may conclude that α_{Jt} is likely different from zero. On the other hand, if most of $\hat{\alpha}_{Jt}$ and $\hat{\alpha}_{Jt,-h}$ are below 95 or 97.5% of $\tilde{\alpha}_{it,-h}$, we conclude that α_{Jt} may not be different from zero.[16]

In practice, instead of using a placebo estimate $\tilde{\alpha}_{it,-h}$ (and baseline and leave-one-out estimates $\hat{\alpha}_{Jt}$ and $\hat{\alpha}_{Jt,-h}$) in each year ($t > T_0$), we use the following three placebo SC estimates (and their counterpart SC estimates) as test statistics in our placebo tests.

Note, however, that this placebo test is clearly different from both the original Fisher's randomization inference (Fisher 1937) and its extension to non-randomized observational studies (Rosenbaum 1984; 2002; Ho and Imai 2006) in the sense that we cannot argue that the "label" of treatment or intervention status is randomized in this placebo test, even after conditioning on covariates.

[15] Note that the real treated units (i.e. Japan) are omitted from the donor pool in iterated placebo SC estimation for easier interpretation of placebo estimates.

[16] The threshold of 95% or 97.5% is based on the counterpart threshold of 90% or 95% significance level in normal two-sided tests. Note, however, that our placebo test is still informal as is the original.

$$\begin{aligned}\text{Test statistic 1}: &\sum_{t>T_0} \tilde{\alpha}_{it,-h} \\ \text{Test statistic 2}: &\sum_{t>T_0} \tilde{\alpha}_{it,-h} - \sum_{t\leq T_0} \tilde{\alpha}_{it,-h} \\ \text{Test statistic 3}: &\sum_{t>T_0} \tilde{\alpha}_{it,-h} - \tilde{\alpha}_{iT_0,-h}\end{aligned} \tag{5}$$

Test statistic 1 is the baseline statistic in which placebo SC estimates are simply averaged over post-intervention year s. Test statistic 2 is the gap between an average post-intervention SC estimate (i.e. test statistic 1) and an average pre-intervention SC estimate. An intuitive rationale for this test statistic is that the two deviation terms $\boldsymbol{\theta}_t(\mathbf{Z}_i - \sum_{k\neq i} w_k^* \mathbf{Z}_k)$ and $\boldsymbol{\lambda}_t(\boldsymbol{\mu}_i - \sum_{k\neq i} w_k^* \boldsymbol{\mu}_k)$ in equation (4) should be similar between pre and post-intervention periods if the coefficients $\boldsymbol{\theta}_t$ and $\boldsymbol{\lambda}_t$ do not change much over time. If this is the case, subtracting the average pre-intervention SC estimate from the average post-intervention SC estimate may reduce some biases caused by insufficient balance in predictors.[17]

The idea of test statistic 3 is similar to that of test statistic 2. In this statistic, the pre-intervention SC estimate at $t = T_0$ is subtracted from test statistic 1 instead of subtracting the average pre-intervention estimate. This test statistic may be better than test statistic 2 in some cases because SC estimates at $t = T_0$ may reflect some bias (i.e. a poor pre-intervention fit) that arises just before the intervention and does not disappear after the intervention. In all three cases we "test" the test statistics for Japan by comparing these values to placebo distributions based on equation (5).

Finally, Abadie et al. (2010) and subsequent studies use the post/pre-intervention ratios of MSPE (mean squared prediction error) to evaluate SC estimates relative to their placebo counterparts. We also use the MSPE ratio of each placebo trial as an additional test statistic:

$$\text{Test statistic 4}: \frac{\sum_{t>T_0} \tilde{\alpha}^2_{it,-h}}{\sum_{t\leq T_0} \tilde{\alpha}^2_{it,-h}} \tag{6}$$

6.3 Results

Figure 6 shows the results of the extended placebo tests for in-kind benefits for the elderly. The distributions of the four test statistics based on formulas (5) and (6) are shown as cumulative distribution functions (CDFs) and the values of baseline and leave-one-out

[17] For example, if we assume that $t = pre$ and $post$, $\alpha_{i,pre} = 0$, and θ_t and λ_t are constant over time, then $\hat{\alpha}_{i,post} - \hat{\alpha}_{i,pre} = \alpha_{i,post} + (\varepsilon_{i,post} - \varepsilon_{i,pre}) - \sum_{j\neq i} w_j^*(\varepsilon_{j,post} - \varepsilon_{j,pre})$.

test statistics for Japan are presented as large and small circles respectively.[18]

All the graphs in Figure 6 show that the baseline and leave-one-out test statistics for Japan are well above most of the placebo-based statistics. Although some placebo-based test statistics are higher than the baseline and leave-one-out test statistics for Japan when test statistics 1 and 2 are used, the third graph suggests that if test statistic 3 is used the baseline statistic and all leave-one-out test statistics for Japan are higher than any other placebo-based statistics due to significant decreases in some highest-value placebo estimates. This implies that these highest-value placebo estimates may have significant upward bias due to an unsuccessful pre-intervention fit in one year before the placebo LTCI introduction.[19] The last graph also shows that the post/pre-intervention MSPE ratios for baseline and leave-one-out SC estimates are higher than any placebo counterparts.

Figures 7 and 8 provide our extended placebo results for the female LFP rates of ages 50-54 and ages 55-59 respectively. Figure 7 shows that the values of test statistics for Japan are in the middle of the placebo-based test statistics, implying that they are not significantly different from the counterpart placebo-based test statistics.

Figure 8 suggests that the baseline test statistics for Japan may be significantly different from the placebo-based test statistics, supporting the placebo result of the LFP rate for ages 55-59 in Figure 5. Leave-one-out test statistics for Japan are however unstable, implying the baseline SC estimates are imprecise.

Finally, placebo results for the other outcomes show that the values of Japan's baseline and most leave-one-out test statistics for these outcomes are not significantly different from the counterpart placebo-based test statistics (online Appendix).

Overall, our extended placebo tests indicate that there exist positive LTCI effects on in-kind LTC benefits for the elderly, but neither positive nor negative LTCI impacts on female LFP rates are robustly observed. At the same time, our placebo tests also reveal the possibility that the female LFP rates for ages 55-59 stagnated after LTCI introduction, although we cannot confirm LTCI introduction directly caused this stagnation.

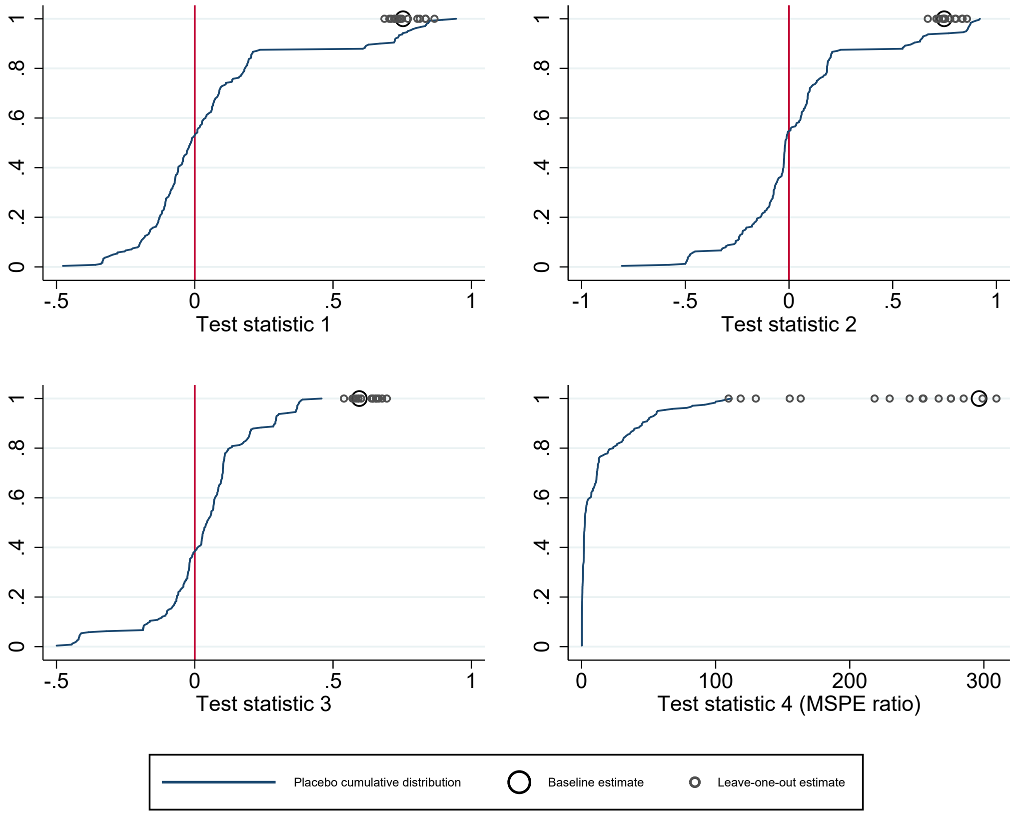
7 Discussion and conclusion

The nationwide LTCI introduction in Japan is one of the major social welfare reforms carried out in the 1990s and 2000s in aging OCED countries. In this paper we investigate the impact of this LTCI introduction on fiscal outcomes and female labor force participation, exploiting the quasi-experimental features of LTCI introduction and using

[18] Note that `nested` and `allopt` options in `synth` command in Stata are not used in all estimations, including baseline and leave-one-out estimations for Japan, to avoid any optimization errors and shorten the computation time.

[19] In fact, the highest placebo values in test statistics 1 and 2 come from the SC estimates of Australia, the pre-intervention fits of which tend to be very poor just before the year 2000.

Figure 6: Extended placebo tests for in-kind benefits for the elderly



Notes: The definitions of test statistics 1-4 are based on equations (5) and (6). In each graph, the X axis shows the values of test statistics and the Y axis indicates the cumulative probability density of placebo-based test statistics. The large circle shows the value of a baseline test statistic for Japan and small circles show the values of leave-one-out test statistics for Japan. Note that the `nested` and `allopt` options in the `synth` command in Stata are not used in all estimations, including baseline and leave-one-out estimations for Japan.

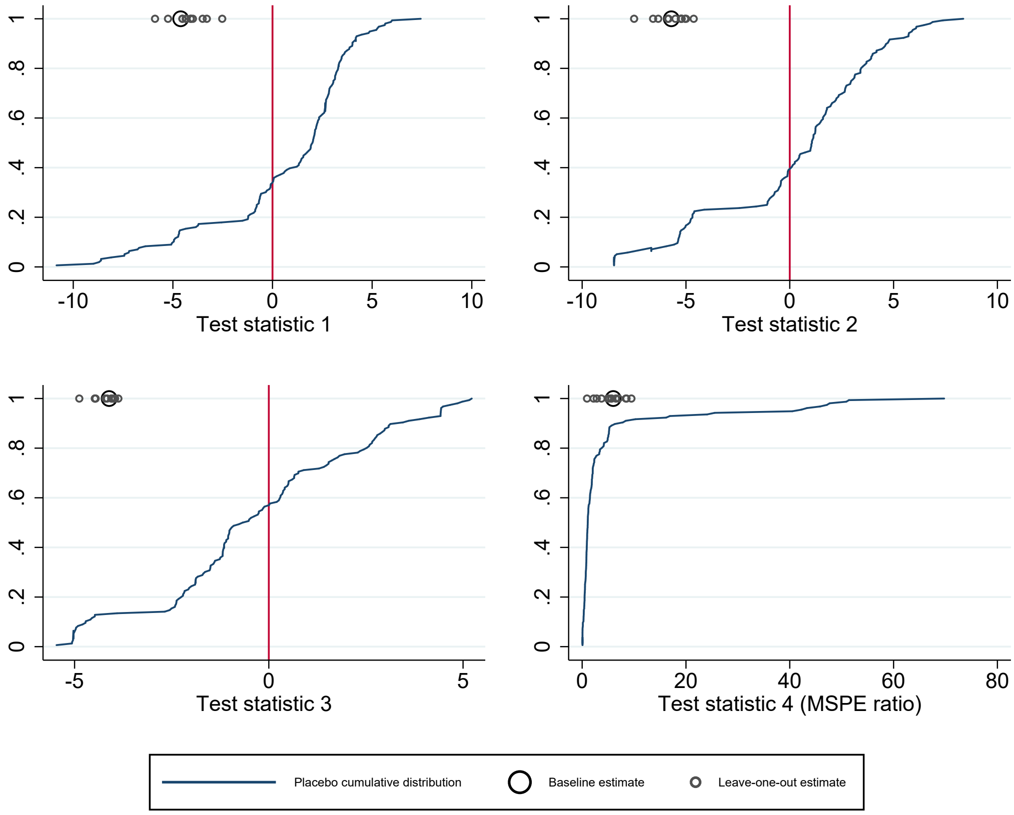
a synthetic control method.

Our estimation results imply that LTCI introduction had a positive impact on the target expenditure item in Japan (i.e. in-kind benefits for the elderly) but we did not find robust effects on public health expenditure or female labor force participation rates. These results suggest that the LTCI program in Japan has not played a sufficient role to alter the family-dependent character of long-term care provision and low female labor force participation in this country.

This macro-level finding in our study may not be consistent with several recent micro-level studies that found some positive labor-supply effects of LTCI's in-kind benefits. Given the fact that we estimate aggregate LTCI effects whereas the previous studies study individual-level LTCI effects, however, we can provide several possible explanations that are consistent with both findings.

First of all, it is possible that we failed to detect some positive LTCI effects on female LFP rates. The power of our SC estimation may not be high enough and there is also

Figure 7: Extended placebo tests for the female LFP rate of ages 50–54



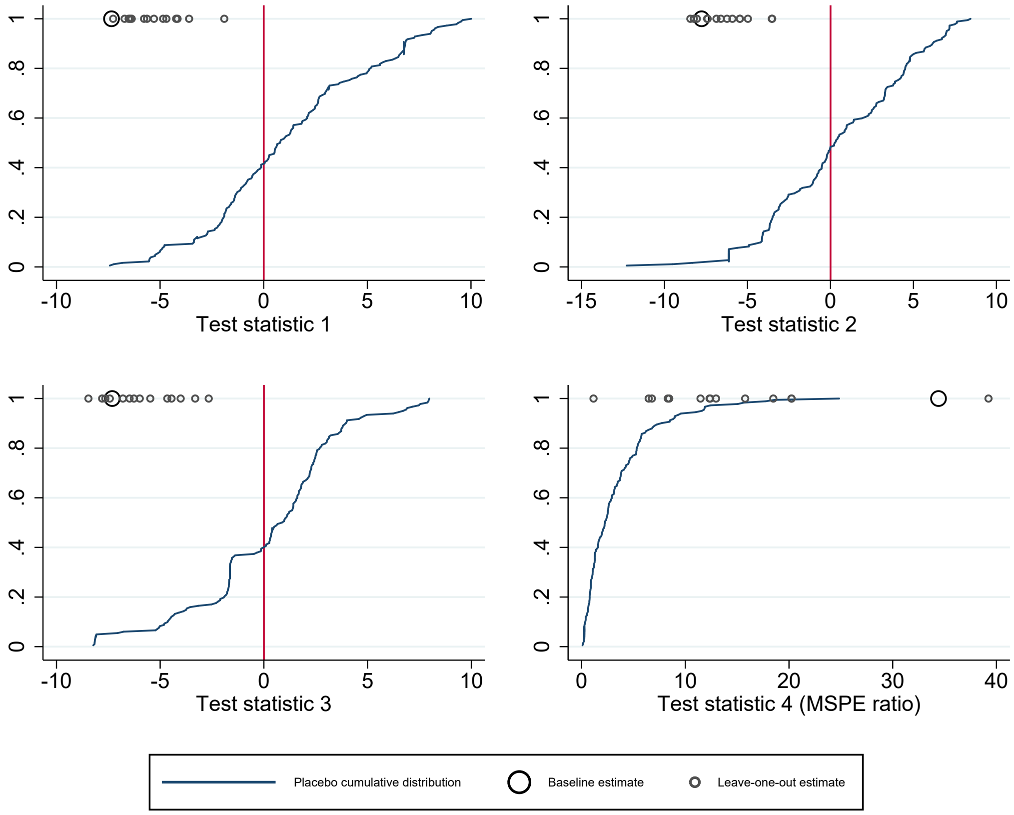
Notes: The definitions of test statistics 1-4 are based on equations (5) and (6). In each graph, the X axis shows the values of test statistics and the Y axis indicates the cumulative probability density of placebo-based test statistics. The large circle shows the value of a baseline test statistic for Japan and small circles show the values of leave-one-out test statistics for Japan. Note that the `nested` and `allopt` options in the `synth` command in Stata are not used in all estimations, including baseline and leave-one-out estimations for Japan.

a possibility that we failed to eliminate some Japan-specific negative confounding trends in the female labor supply. Even if this is the case, however, our analysis and several robustness checks still imply that an aggregate positive LTCI effect on female labor force participation, if it exists, is small enough to remain undetected by our analysis.

Second, LTCI benefits may have enabled more frail elderly people to live at home with their family. If this is the case, it is possible that some family caregivers worked more because of more in-kind benefits from LTCI (i.e. a positive effect), but some people worked less because they chose to be family caregivers for elderly people who would have been in hospitals or nursing homes if LTCI had not been introduced (i.e. a negative effect). Most individual-level studies focus on the first effect, but our aggregate-level study is meant to capture both effects.[20]

[20]This cancelling-out negative effect is at least somewhat plausible given the fact that Japanese LTCI has mostly led to increases in residential care rather than institutional care. The ratio of the elderly who received residential LTCI services increased from 4.4 % in 2000 to 12.4 % in 2015. On the other hand, the total capacity of institutional care for the elderly (both public and private) only increased

Figure 8: Extended placebo tests for the female LFP rate of ages 55–59



Notes: The definitions of test statistics 1-4 are based on equations (5) and (6). In each graph, the X axis shows the values of test statistics and the Y axis indicates the cumulative probability density of placebo-based test statistics. The large circle shows the value of a baseline test statistic for Japan and small circles show the values of leave-one-out test statistics for Japan. Note that the `nested` and `allopt` options in the `synth` command in Stata is not used in all estimations, including baseline and leave-one-out estimations for Japan.

Third, there may be other negative LTCI effects on female LFP rates such as an income effect caused by this new public insurance scheme and changes in female labor supply and demand caused by new insurance premiums for individuals and firms. We were not able to address these issues in this paper, but it seems unlikely that these possible negative effects are large enough to cancel out the expected positive effects.

Our study only revealed that the Japanese LTCI introduction clearly boosted LTC spending but failed to boost labor force participation for middle-aged women. We discussed some possible mechanisms behind these results, but the mystery of no aggregate

from 3.7 % in 2000 to 5.5 % in 2015. In addition, the number of long-term elderly inpatients (including social hospitalization) significantly decreased after the introduction of LTCI in 2000: the ratio of the elderly who were hospitalized for more than one month decreased from 2.5 % in 1999 to 1.7 % in 2014. These statistics imply that more old people who need health and social care now stay at home for a longer period using formal LTC services. This is exactly what the Japanese government intended to achieve through LTCI (Campbell and Ikegami 2000; 2003), but this may increase the burden on some informal caregivers who would not have become caregivers if the elderly they take care of had instead been admitted to hospitals for a long period or stayed in nursing homes.

positive LTCI effect remains. Further studies are required to address this question and reconsider the potential roles of LTCI in female labor force participation.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010) "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American statistical Association*, Vol. 105, No. 490, pp. 493–505.

——— (2015) "Comparative politics and the synthetic control method," *American Journal of Political Science*, Vol. 59, No. 2, pp. 495–510.

Abadie, Alberto and Javier Gardeazabal (2003) "The economic costs of conflict: A case study of the Basque Country," *American Economic Review*, Vol. 93, No. 1, pp. 113–132.

Alesina, Alberto and Paola Giuliano (2010) "The power of the family," *Journal of Economic Growth*, Vol. 15, No. 2, pp. 93–125.

Arnold, Tobias and Isabelle Stadelmann-Steffen (2017) "How federalism influences welfare spending: Belgium federalism reform through the perspective of the synthetic control method," *European Journal of Political Research*, Vol. 56, No. 3, pp. 680–702.

Barczyk, Daniel and Matthias Kredler (2017) "Evaluating long-term-care policy options, taking the family seriously," *The Review of Economic Studies*, Vol. 85, No. 2, pp. 766–809.

——— (2018) "Long-term care across Europe and the US: The role of informal and formal care."

Barlow, Pepita (2018) "Does trade liberalization reduce child mortality in low-and middle-income countries? A synthetic control analysis of 36 policy experiments, 1963-2005," *Social Science & Medicine*, Vol. 205, pp. 107–115.

Bauer, Jan Michael and Alfonso Sousa-Poza (2015) "Impacts of informal caregiving on caregiver employment, health, and family," *Journal of Population Ageing*, Vol. 8, No. 3, pp. 113–145.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and John Reenen (2004) "Evaluating the employment impact of a mandatory job search program," *Journal of the European Economic Association*, Vol. 2, No. 4, pp. 569–606.

Bolin, Kristian, Bjorn Lindgren, and Petter Lundborg (2008a) "Informal and formal care among single-living elderly in Europe," *Health Economics*, Vol. 17, No. 3, pp. 393–409.

Bolin, Kristian, Björn Lindgren, and Petter Lundborg (2008b) "Your next of kin or your own career?: Caring and working among the 50+ of Europe," *Journal of Health Economics*, Vol. 27, No. 3, pp. 718–738.

Bonsang, Eric (2009) "Does informal care from children to their elderly parents substitute for formal care in Europe?" *Journal of Health Economics*, Vol. 28, No. 1, pp. 143–154.

Brady, David, Evelyne Huber, and John D. Stephens (2014) *Comparative Welfare States Data Set*: University of North Carolina and WZB Berlin Social Science Center.

Campbell, John Creighton (2002) "How policies differ: long-term-care insurance in Japan and Germany," in Harald Conrad and Ralph Lutzeler eds. *Aging and Social Policy—A German-Japanese Comparison*: Iudicium 2002, pp. 157–187.

Campbell, John Creighton and Naoki Ikegami (2000) "Long-term care insurance comes to Japan," *Health Affairs*, Vol. 19, No. 3, pp. 26–39.

——— (2003) "Japan's Radical Reform of Long-term Care," *Social Policy & Administration*, Vol. 37, No. 1, pp. 21–34.

Campbell, John Creighton, Naoki Ikegami, and Mary Jo Gibson (2010) "Lessons from public long-term care insurance in Germany and Japan," *Health Affairs*, Vol. 29, No. 1, pp. 87–95.

Carmichael, Fiona and Sue Charles (1998) "The labour market costs of community care1," *Journal of Health Economics*, Vol. 17, No. 6, pp. 747–765.

Carmichael, Fiona and Susan Charles (2003a) "Benefit payments, informal care and female labour supply," *Applied Economics Letters*, Vol. 10, No. 7, pp. 411–415.

——— (2003b) "The opportunity costs of informal care: does gender matter?" *Journal of Health Economics*, Vol. 22, No. 5, pp. 781–803.

Casado-Marín, David, Pilar García-Gómez, and Ángel López-Nicolás (2011) "Informal care and labour force participation among middle-aged women in Spain," *SERIEs*, Vol. 2, No. 1, pp. 1–29.

Charles, Kerwin Kofi and Purvi Sevak (2005) "Can family caregiving substitute for nursing home care?" *Journal of Health Economics*, Vol. 24, No. 6, pp. 1174–1190.

Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu (2018) "An exact and robust conformal inference method for counterfactual and synthetic controls," *arXiv preprint arXiv:1712.09089*.

Ciani, Emanuele (2012) "Informal adult care and caregivers' employment in Europe," *Labour Economics*, Vol. 19, No. 2, pp. 155–164.

Colombo, F, A Llena-Nozal, J Mercier, and T Frits (2011) "Public long-term care financing arrangements in OECD countries," in *Help Wanted?: Providing and Paying for Long-Term Care*: OECD publishing, Chap. 7, pp. 213–246.

Crespo, Laura and Pedro Mira (2014) "Caregiving to elderly parents and employment status of European mature women," *Review of Economics and Statistics*, Vol. 96, No. 4, pp. 693–709.

Doudchenko, Nikolay and Guido W Imbens (2017) "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis," *arXiv preprint arXiv:1610.07748*.

Esping-Andersen, Gøsta (1997) "Hybrid or unique?: The Japanese welfare state between Europe and America," *Journal of European Social Policy*, Vol. 7, No. 3, pp. 179–189.

——— (1999) *Social foundations of postindustrial economies*: Oxford University Press.

Ettner, Susan L. (1995) "The impact of "parent care" on female labor supply decisions," *Demography*, Vol. 32, No. 1, pp. 63–80.

Ettner, Susan L (1996) "The opportunity costs of elder care," *Journal of Human Resources*, pp. 189–205.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer (2015) "The next generation of the Penn World Table," *American Economic Review*, Vol. 105, No. 10, pp. 3150–82.

Ferman, Bruno and Cristine Pinto (2017) "Placebo tests for synthetic controls," *MPRA Paper*, No. 78079.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom (2018) "Cherry picking with synthetic controls," *MPRA Paper Paper*, No. 85138.

Finkelstein, Amy (2007) "The aggregate effects of health insurance: Evidence from the introduction of Medicare," *The Quarterly Journal of Economics*, Vol. 122, No. 1, pp. 1–37.

Firpo, Sergio and Vitor Possebom (2018) "Synthetic control method: Inference, sensitivity analysis and confidence sets," *Journal of Causal Inference*, Vol. 6, No. 2.

Fisher, Ronald Aylmer (1937) *The design of experiments*: Oliver And Boyd; Edinburgh; London.

Fu, Rong, Haruko Noguchi, Akira Kawamura, Hideto Takahashi, and Nanako Tamiya (2017) "Spillover effect of Japanese long-term care insurance as an employment promotion policy for family caregivers," *Journal of Health Economics*, Vol. 56, pp. 103–112.

Fukahori, Ryotaro, Tadashi Sakai, and Kazuma Sato (2015) "The Effects of Incidence of Care Needs in Households on Employment, Subjective Health, and Life Satisfaction among Middle-aged Family Members," *Scottish Journal of Political Economy*, Vol. 62, No. 5, pp. 518–545.

Geyer, Johannes and Thorben Korfhage (2015) "Long-term Care Insurance and Carers' Labor Supply–A Structural Model," *Health Economics*, Vol. 24, No. 9, pp. 1178–1191.

——— (2018) "Labor supply effects of long-term care reform in Germany," *Health Economics*, Vol. 27, No. 9, pp. 1328–1339.

Goldin, Claudia (2006) "The quiet revolution that transformed women's employment, education, and family," *American Economic Review*, Vol. 96, No. 2, pp. 1–21.

——— (2014) "A grand gender convergence: Its last chapter," *American Economic Review*, Vol. 104, No. 4, pp. 1091–1119.

Hanaoka, Chie and Edward C. Norton (2008) "Informal and formal care for elderly persons: How adult children's characteristics affect the use of formal care in Japan," *Social Science & Medicine*, Vol. 67, No. 6, pp. 1002 – 1008.

Heckman, James J, Lance Lochner, and Christopher Taber (1998) "General equilibrium treatment effects: A study of tuition policy,"Technical report, National Bureau of Economic Research.

Heitmueller, Axel (2007) "The chicken or the egg?: Endogeneity in labour market participation of informal carers in England," *Journal of Health Economics*, Vol. 26, No. 3, pp. 536–559.

Heitmueller, Axel and Kirsty Inglis (2007) "The earnings of informal carers: Wage differentials and opportunity costs," *Journal of Health Economics*, Vol. 26, No. 4, pp. 821 – 841.

Hieda, Takeshi (2012) *Political Institutions and Elderly Care Policy: Comparative politics of long-term care in advanced democracies*: Springer.

Ho, Daniel E and Kosuke Imai (2006) "Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election," *Journal of the American Statistical Association*, Vol. 101, No. 475, pp. 888–900.

Johnson, Richard W and Anthony T Lo Sasso (2006) "The impact of elder care on women's labor supply," *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, Vol. 43, No. 3, pp. 195–210.

Kondo, Ayako (2017) "Availability of Long-term Care Facilities and Middle-aged People's Labor Supply in Japan," *Asian Economic Policy Review*, Vol. 12, No. 1, pp. 95–112.

Kotsadam, Andreas (2011) "Does Informal Eldercare Impede Women's Employment? The Case of European Welfare States," *Feminist Economics*, Vol. 17, No. 2, pp. 121–144.

Leigh, Andrew (2010) "Informal care and labor market participation," *Labour Economics*, Vol. 17, No. 1, pp. 140–149.

Lilly, Meredith B, Audrey Laporte, and Peter C Coyte (2007) "Labor market work and home care's unpaid caregivers: a systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work," *The Milbank Quarterly*, Vol. 85, No. 4, pp. 641–690.

Lilly, Meredith B., Audrey Laporte, and Peter C. Coyte (2010) "Do they care too much to work? The influence of caregiving intensity on the labour force participation of unpaid caregivers in Canada," *Journal of Health Economics*, Vol. 29, No. 6, pp. 895 – 903.

Løken, Katrine V, Shelly Lundberg, and Julie Riise (2016) "Lifting the Burden: Formal Care of the Elderly and Labor Supply of Adult Children," *Journal of Human Resources*, pp. 0614–6447R1.

Mandel, Hadas and Moshe Semyonov (2006) "A welfare state paradox: State interventions and women's employment opportunities in 22 countries," *American Journal of Sociology*, Vol. 111, No. 6, pp. 1910–1949.

Meng, Annika (2012) "Informal caregiving and the retirement decision," *German Economic Review*, Vol. 13, No. 3, pp. 307–330.

——— (2013) "Informal home care and labor-force participation of household members," *Empirical Economics*, Vol. 44, No. 2, pp. 959–979.

Michaud, Pierre-Carl, Axel Heitmueller, and Zafar Nazarov (2010) "A dynamic analysis of informal care and employment in England," *Labour Economics*, Vol. 17, No. 3, pp. 455–465.

Moscarola, Flavia Coda (2010) "Informal caregiving and women's work choices: lessons from the Netherlands," *Labour*, Vol. 24, No. 1, pp. 93–105.

Ngai, L Rachel and Christopher A Pissarides (2011) "Taxes, social subsidies, and the allocation of work time," *American Economic Journal: Macroeconomics*, Vol. 3, No. 4, pp. 1–26.

Nguyen, Ha Trong and Luke Brian Connelly (2014) "The effect of unpaid caregiving intensity on labour force participation: Results from a multinomial endogenous treatment model," *Social Science & Medicine*, Vol. 100, pp. 115–122.

Obama, Barack (2016) "United States health care reform: progress to date and next steps," *JAMA.*, Vol. 316, No. 5, pp. 525–532.

OECD (2012) *Closing the gender gap: Act now*: OECD Paris.

——— (2017) *The Pursuit of Gender Equality*: OECD Paris, pp.304.

Olivetti, Claudia and Barbara Petrongolo (2016) "The evolution of gender gaps in industrialized countries," *Annual Review of Economics*, Vol. 8, pp. 405–434.

——— (2017) "The economic consequences of family policies: lessons from a century of legislation in high-income countries," *Journal of Economic Perspectives*, Vol. 31, No. 1, pp. 205–30.

Olper, Alessandro, Daniele Curzi, and Johan Swinnen (2018) "Trade liberalization and child mortality: a synthetic control method," *World Development*, Vol. 110, pp. 394–410.

Oshio, Takashi and Emiko Usui (2017) "Informal parental care and female labour supply in Japan," *Applied Economics Letters*, Vol. 24, No. 9, pp. 635–638.

——— (2018) "How does informal caregiving affect daughters' employment and mental health in Japan?" *Journal of the Japanese and International Economies*, Vol. 49, pp. 1–7.

Podestà, Federico (2017) "The impact of 'free choice': Family reforms of France and Belgium, a synthetic control analysis," *International Journal of Social Welfare*, Vol. 26, No. 4, pp. 340–352.

Restrepo, Brandon J and Matthias Rieger (2016) "Denmark's policy on artificial trans fat and cardiovascular disease," *American journal of Preventive Medicine*, Vol. 50, No. 1, pp. 69–76.

Rhee, Jong Chul, Nicolae Done, and Gerard F Anderson (2015) "Considering long-term care insurance for middle-income countries: comparing South Korea with Japan and Germany," *Health Policy*, Vol. 119, No. 10, pp. 1319–1329.

Rieger, Matthias, Natascha Wagner, and Arjun S Bedi (2017) "Universal health coverage at the macro level: synthetic control evidence from Thailand," *Social Science & Medicine*, Vol. 172, pp. 46–55.

Rosenbaum, Paul R (1984) "Conditional permutation tests and the propensity score in observational studies," *Journal of the American Statistical Association*, Vol. 79, No. 387, pp. 565–574.

——— (2002) "Observational studies," in *Observational studies*: Springer, pp. 1–17.

Ryan, Andrew M, Sam Krinsky, Evangelos Kontopantelis, and Tim Doran (2016) "Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study," *The Lancet*, Vol. 388, No. 10041, pp. 268–274.

Schmitz, Hendrik and Matthias Westphal (2017) "Informal care and long-term labor market outcomes," *Journal of Health Economics*, Vol. 56, pp. 1–18.

Shimizutani, Satoshi, Wataru Suzuki, and Haruko Noguchi (2008) "The socialization of at-home elderly care and female labor market participation: Micro-level evidence from Japan," *Japan and the World Economy*, Vol. 20, No. 1, pp. 82–96.

Skira, Meghan M (2015) "Dynamic wage and employment effects of elder parent care," *International Economic Review*, Vol. 56, No. 1, pp. 63–93.

Sommers, Benjamin D, Atul A Gawande, and Katherine Baicker (2017) "health Insurance Coverage and Health—What the Recent Evidence Tells Us," *The New England Journal of Medicine*, Vol. 377, No. 6, pp. 586–593.

Sugawara, Shinya and Jiro Nakamura (2014) "Can formal elderly care stimulate female labor supply? The Japanese experience," *Journal of the Japanese and International Economies*, Vol. 34, pp. 98–115.

Tamiya, Nanako, Haruko Noguchi, Akihiro Nishi, Michael R Reich, Naoki Ikegami, Hideki Hashimoto, Kenji Shibuya, Ichiro Kawachi, and John Creighton Campbell (2011) "Population ageing and wellbeing: lessons from Japan's long-term care insurance policy," *The Lancet*, Vol. 378, No. 9797, pp. 1183–1192.

Van Houtven, Courtney Harold, Norma B Coe, and Meghan M Skira (2013) "The effect of informal care on work and wages," *Journal of Health Economics*, Vol. 32, No. 1, pp. 240–252.

Van Houtven, Courtney Harold and Edward C Norton (2004) "Informal care and health care use of older adults," *Journal of Health Economics*, Vol. 23, No. 6, pp. 1159–1180.

Wolf, Douglas A and Beth J Soldo (1994) "Married women's allocation of time to employment and care of elderly parents," *Journal of Human Resources*, pp. 1259–1276.

Yamada, Hiroyuki and Satoshi Shimizutani (2015) "Labor market outcomes of informal care provision in Japan," *The Journal of the Economics of Ageing*, Vol. 6, pp. 79–88.

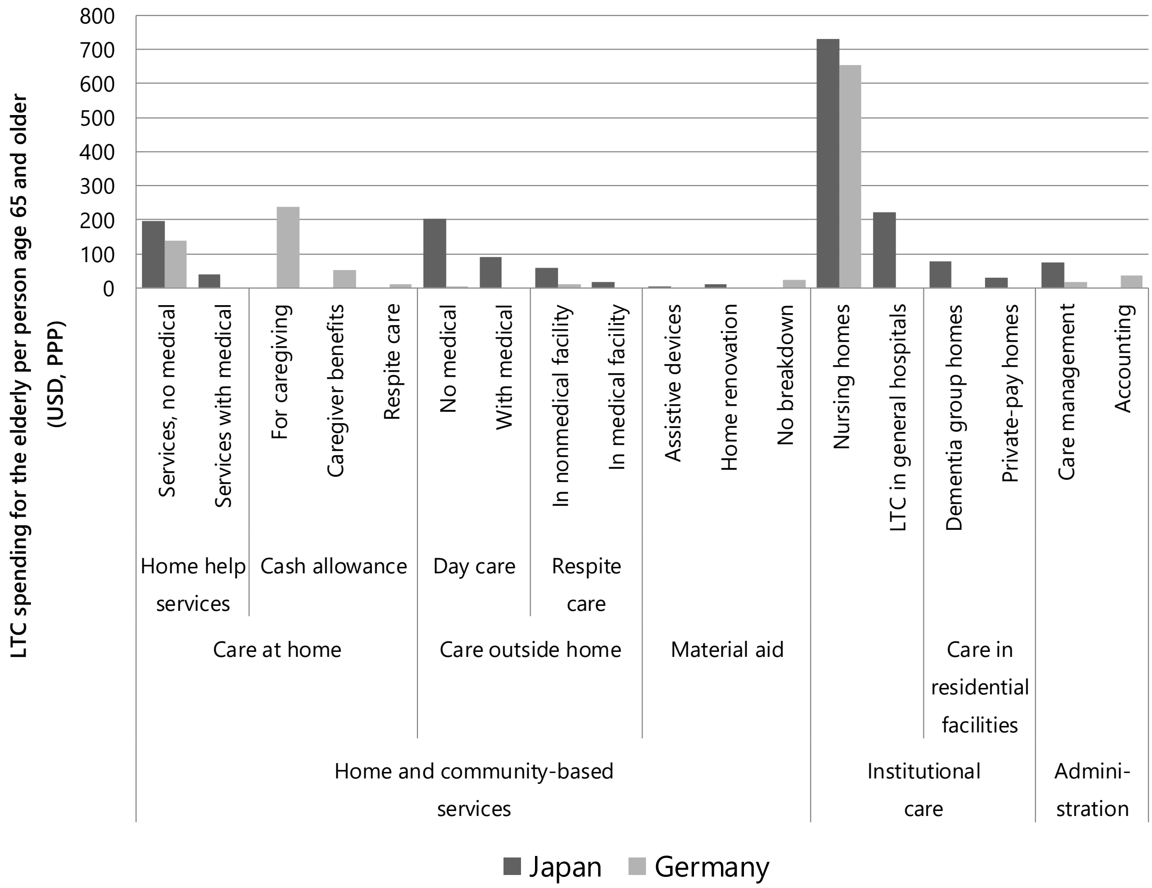
Appendices

A Comparisons of LTCI in Japan and Germany

Table A.1: Institutional settings of LTCI programs in Germany and Japan, 2008 (Campbell et al. 2010)

	Germany	Japan
Policy objective	Support family caregivers	Decrease burden of family caregivers
Policy design	Contain spending to within the premium level set by law	Increase expenditures as services become more available
Organized and managed	Sickness funds (but LTC is managed separately)	LTC insurance section of the municipal government or their coalitions
Financing	Premiums	Half by premiums, half by taxes
	1.95% of income up to a ceiling	1/3 of premium revenue from those age 65+, with 6 premium levels based on income
		2/3 from those ages 40–64 at 1% of income, up to a ceiling
Regional differences	No difference in premium levels	For those age 65+, premiums linked to local spending level
		For those ages 40–64, pooled at national level and redistributed; municipalities having low income levels and more residents age 75+ receive more
Population covered	All ages	Unconditional for those age 65+
		Limited to age-related diseases for those ages 40–64
Percentage of those age 65+ who are eligible	10.50%	17%
Percentage of those age 65+ who are receiving benefits	10.5% (all of those eligible receive benefits)	13.5% (20% of those eligible have not chosen to receive benefits)
Eligibility levels	Three (plus a limited additional hardship level in HCBS)	Two for the "preventive care" program, five for regular LTC insurance
Benefit ceilings per month ($ PPP)	No copayment or deductible	10% copayment (included below)
	Cash: $250–$794 plus caregiver pension premiums	Services only
	HCBS: $490–$1,730 (hardship: $2,260)	HCBS preventive care: $430–$950
	Institutional care: $1,200–$1,730	HCBS regular program: $1,440–$3,400
	Room-and-board costs not covered; low income covered under public assistance	Institutional care: $1,680–$3,670
		One-third of room-and-board costs covered, up to all costs paid by LTC insurance if income level is low
Fee schedule for services	Negotiated regionally between sickness funds and providers	Negotiated nationally, conversion factor for regional cost differences

Table A.2: LTC spending on the elderly per person aged 65 and older (USD, PPP, 2005)



Source: Exibit 2 of Campbell et al. (2010).

B Variable definitions and sources

Table B.1: Variable definitions and sources

Outcome	Definition	Source
Public expenditure on old-age benefits in kind	Public expenditure on old-age benefits in kind as percentages of GDP	OECD Database
Total public expenditure on health care	Total public expenditure on health care, as a percentage of GDP	OECD Database
Total public social expenditure	Total public social expenditure, as percentages of GDP	OECD Database
Female labour force participation rate by age cohort	Ratio of women in the labor force in an age cohort to the female population of the same age cohort	OECD Database

Predictor	Definition	Source
Employment in agriculture	Employment in agriculture as % of civilian employment	CWS
Employment in industry	Employment in industry as % of civilian employment	CWS
Employment in services	Employment in services as % of civilian employment	CWS
Child population	Population under 15 as % population.	OECD Database
Elderly population	Population 65 and over as % population	OECD Database
Growth rate of chilld population	Growth rate of child population calculated by authors	OECD Database
Growth rate of elderly population	Growth rate of Elderly population calculated by authors	OECD Database
Per capita real GDP	Expenditure-siderealGDP atchained PPPs(in mil. 2005US$) divided by population	PWT8.1
Growth rate of real GDP	calculated by authors	PWT8.1

Notes: CWS and PWT 8.1 indicate "Comparative Welfare State Dataset" (Brady et al. 2014) and "Penn World Table 8,1" (Feenstra et al. 2015) respectively. Expenditure data based on OECD database are from OECD Social Expenditure Database (SOCX). Demographic data based on OECD databases are from OECD Employment and Labour Force Statistics.

Online Appendices

I Weights and predictor balance

Table I.1: Outcome: in-kind benefits for the elderly

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0.064	0.087	0.078
Austria	0.525	-	-
Belgium	0.186	0	0
Denmark	0	0.016	0.002
Finland	0	0	0
France	0	0	0
Ireland	0	0	0.048
Italy	0	0	0.787
NewZealand	0	0	0
Norway	0	0	0
Portugal	0	0.855	-
Spain	0.222	0	0
Sweden	0.002	0.042	0.033
Switzerland	0	0	0
UnitedKingdom	0	0	0.052
UnitedStates	0	0	0

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1980)	0.102	0.112	0.093	0.093
Lagged outcome (1981)	0.105	0.112	0.096	0.095
Lagged outcome (1982)	0.111	0.113	0.096	0.095
Lagged outcome (1983)	0.113	0.112	0.095	0.096
Lagged outcome (1984)	0.111	0.112	0.092	0.094
Lagged outcome (1985)	0.110	0.112	0.094	0.095
Lagged outcome (1986)	0.124	0.109	0.095	0.097
Lagged outcome (1987)	0.117	0.133	0.130	0.128
Lagged outcome (1988)	0.116	0.133	0.137	0.131
Lagged outcome (1989)	0.121	0.134	0.139	0.129
Lagged outcome (1990)	0.128	0.158	0.169	0.175
Lagged outcome (1991)	0.139	0.166	0.182	0.184
Lagged outcome (1992)	0.155	0.159	0.202	0.200
Lagged outcome (1993)	0.169	0.165	0.216	0.208
Lagged outcome (1994)	0.185	0.183	0.208	0.211
Lagged outcome (1995)	0.219	0.199	0.209	0.210
Lagged outcome (1996)	0.227	0.230	0.214	0.219
Lagged outcome (1997)	0.247	0.229	0.219	0.225
Lagged outcome (1998)	0.192	0.279	0.274	0.271
Lagged outcome (1999)	0.444	0.291	0.272	0.261
GDP per capita	22.805	20.994	14.182	21.724
Share of children (%)	18.919	18.955	20.929	18.395
Share of the elderly (%)	12.294	14.337	13.358	14.604
Share of agriculture (%)	7.409	8.026	16.485	8.735
Share of industry (%)	34.082	33.061	33.253	32.738
Share of service (%)	58.509	58.796	50.239	58.506
GDP per capita (growth rate)	2.827	2.980	3.890	2.845
Population (growth rate)	0.436	0.380	0.354	0.256
Share of children (growth rate)	-2.351	-1.333	-2.065	-2.067
Share of the elderly (growth rate)	3.219	0.663	1.716	1.391

Table I.2: Outcome: public health expenditure

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0	0.129	0.335
Austria	0.018	0	0.295
Belgium	0	0	0
Denmark	0	0	0
Finland	0	0.065	0
France	0	0	0
Ireland	0.297	0.301	0.119
Italy	0.174	0.182	0.133
NewZealand	0	0	0
Norway	0	0	0
Portugal	0.193	0.323	-
Spain	0	0	0
Sweden	0	0	0
Switzerland	0.317	-	-
UnitedKingdom	0	0	0
UnitedStates	0	0	0.117

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1980)	4.446	4.469	4.556	4.518
Lagged outcome (1981)	4.504	4.591	4.660	4.444
Lagged outcome (1982)	4.626	4.526	4.522	4.443
Lagged outcome (1983)	4.790	4.537	4.441	4.494
Lagged outcome (1984)	4.640	4.499	4.487	4.632
Lagged outcome (1985)	4.578	4.588	4.610	4.678
Lagged outcome (1986)	4.662	4.709	4.723	4.752
Lagged outcome (1987)	4.707	4.720	4.678	4.765
Lagged outcome (1988)	4.587	4.731	4.691	4.666
Lagged outcome (1989)	4.486	4.682	4.551	4.648
Lagged outcome (1990)	4.442	4.497	4.645	4.744
Lagged outcome (1991)	4.556	4.743	4.835	4.933
Lagged outcome (1992)	4.769	4.895	4.906	5.102
Lagged outcome (1993)	5.047	5.097	5.149	5.257
Lagged outcome (1994)	5.220	5.066	5.095	5.302
Lagged outcome (1995)	5.567	5.554	5.558	5.443
Lagged outcome (1996)	5.687	5.682	5.653	5.432
Lagged outcome (1997)	5.621	5.665	5.639	5.551
Lagged outcome (1998)	5.784	5.654	5.644	5.590
Lagged outcome (1999)	5.986	5.735	5.791	5.691
GDP per capita	22.805	22.207	18.615	23.419
Share of children (%)	18.919	20.754	21.914	21.502
Share of the elderly (%)	12.294	13.581	13.024	12.688
Share of agriculture (%)	7.409	9.514	11.348	7.197
Share of industry (%)	34.082	30.835	30.020	29.430
Share of service (%)	58.509	59.551	58.598	63.184
GDP per capita (growth rate)	2.827	2.743	3.201	2.935
Population (growth rate)	0.436	0.485	0.502	0.802
Share of children (growth rate)	-2.351	-1.235	-1.476	-0.979
Share of the elderly (growth rate)	3.219	0.763	1.032	0.629

Table I.3: Outcome: female labor force participation rate (ages 30-34)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0	0	0.97
Belgium	0	0	0
Canada	0	0	0
Denmark	0	0	0
Finland	0	0	0
France	0	0	0
Italy	0.4	1	-
NewZealand	0	0	0
Norway	0	0	0
Portugal	0	0	0.03
Spain	0.6	-	-
Sweden	0	0	0
UnitedKingdom	0	0	0
UnitedStates	0	0	0

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	50.000	48.842	58.127	59.267
Lagged outcome (1987)	50.485	52.821	59.234	60.863
Lagged outcome (1988)	50.877	54.838	60.983	60.693
Lagged outcome (1989)	51.145	56.726	62.277	62.842
Lagged outcome (1990)	51.680	58.387	62.821	64.036
Lagged outcome (1991)	52.865	59.799	62.600	63.562
Lagged outcome (1992)	52.727	60.253	63.111	63.867
Lagged outcome (1993)	52.713	60.237	58.925	63.255
Lagged outcome (1994)	53.470	60.542	58.497	64.075
Lagged outcome (1995)	53.652	61.781	59.494	65.892
Lagged outcome (1996)	54.798	62.751	60.643	65.344
Lagged outcome (1997)	56.188	63.891	61.296	65.636
Lagged outcome (1998)	55.769	64.787	62.126	65.797
Lagged outcome (1999)	56.674	66.148	63.326	65.142
GDP per capita	25.479	19.778	23.614	25.874
Share of children (%)	17.315	17.376	15.840	21.741
Share of the elderly (%)	13.419	14.912	15.769	11.575
Share of agriculture (%)	6.511	9.555	7.963	5.633
Share of industry (%)	33.692	32.188	33.083	24.167
Share of service (%)	59.796	58.257	58.954	69.996
GDP per capita (growth rate)	4.126	4.740	3.668	3.006
Population (growth rate)	0.326	0.195	0.073	1.267
Share of children (growth rate)	-2.637	-2.741	-2.153	-0.914
Share of the elderly (growth rate)	3.521	2.280	2.268	1.349

Table I.4: Outcome: female labor force participation rate (ages 35-39)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0	0.72	-
Belgium	0	0	0
Canada	0	0	0
Denmark	0	0	0
France	0	0	0
Italy	0.754	-	-
NewZealand	0.246	0	0.581
Norway	0	0	0
Portugal	0	0	0
Spain	0	0.28	0.419
Sweden	0	0	0
UnitedKingdom	0	0	0
UnitedStates	0	0	0

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	61.002	58.288	55.373	56.287
Lagged outcome (1987)	61.314	60.497	56.951	57.842
Lagged outcome (1988)	61.315	60.991	60.045	58.738
Lagged outcome (1989)	62.370	61.832	61.371	59.908
Lagged outcome (1990)	62.611	61.818	64.233	62.036
Lagged outcome (1991)	62.093	62.882	64.066	64.154
Lagged outcome (1992)	62.379	63.892	64.604	64.117
Lagged outcome (1993)	61.654	60.825	65.010	64.935
Lagged outcome (1994)	61.578	60.854	65.446	64.764
Lagged outcome (1995)	60.465	61.640	66.392	66.181
Lagged outcome (1996)	60.836	62.606	66.940	68.203
Lagged outcome (1997)	62.338	62.298	66.359	67.362
Lagged outcome (1998)	62.211	63.010	65.711	67.124
Lagged outcome (1999)	61.538	64.439	66.481	68.585
GDP per capita	25.479	22.756	23.711	18.909
Share of children (%)	17.315	17.657	20.860	21.203
Share of the elderly (%)	13.419	14.653	12.293	12.536
Share of agriculture (%)	6.511	8.429	6.806	10.175
Share of industry (%)	33.692	30.980	26.031	27.491
Share of service (%)	59.796	60.591	67.012	62.334
GDP per capita (growth rate)	4.126	3.465	3.621	3.937
Population (growth rate)	0.326	0.337	1.016	0.780
Share of children (growth rate)	-2.637	-1.752	-1.500	-1.618
Share of the elderly (growth rate)	3.521	1.938	1.595	1.496

Table I.5: Outcome: female labor force participation rate (ages 40-44)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0.074	0	0.754
Belgium	0	0	0
Canada	0	0	0
Denmark	0	0	0
France	0	0	0
Italy	0.354	0.581	-
NewZealand	0.519	-	-
Norway	0	0	0
Portugal	0	0	0
Spain	0	0	0.103
Sweden	0.053	0.419	0
UnitedKingdom	0	0	0.143
UnitedStates	0	0	0

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	68.793	67.967	66.711	63.126
Lagged outcome (1987)	68.386	68.241	67.130	64.465
Lagged outcome (1988)	68.076	68.506	67.802	66.657
Lagged outcome (1989)	68.750	68.936	69.393	69.099
Lagged outcome (1990)	69.582	69.886	70.122	70.940
Lagged outcome (1991)	70.377	70.019	70.176	71.123
Lagged outcome (1992)	70.513	70.396	70.894	71.455
Lagged outcome (1993)	70.291	68.680	69.979	70.692
Lagged outcome (1994)	69.792	70.023	69.595	71.663
Lagged outcome (1995)	69.469	69.420	69.421	72.359
Lagged outcome (1996)	69.464	69.846	70.239	72.987
Lagged outcome (1997)	70.874	70.048	70.470	72.935
Lagged outcome (1998)	70.175	71.158	71.900	72.357
Lagged outcome (1999)	69.543	71.728	72.060	72.441
GDP per capita	25.479	22.040	23.947	24.747
Share of children (%)	17.315	20.248	16.895	21.091
Share of the elderly (%)	13.419	13.195	16.530	12.396
Share of agriculture (%)	6.511	8.505	6.021	5.403
Share of industry (%)	33.692	27.659	30.670	25.484
Share of service (%)	59.796	63.821	63.309	68.954
GDP per capita (growth rate)	4.126	3.151	3.413	3.273
Population (growth rate)	0.326	0.739	0.220	1.049
Share of children (growth rate)	-2.637	-1.090	-1.185	-0.974
Share of the elderly (growth rate)	3.521	1.383	1.335	1.277

Table I.6: Outcome: female labor force participation rate (ages 45-49)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0	0	0
Belgium	0	0	0.23
Canada	0	0	0
Denmark	0.157	0.315	0
France	0	0	0
Italy	0.378	-	-
NewZealand	0	0	0
Norway	0	0	0
Portugal	0.002	0	0
Spain	0	0.346	-
Sweden	0.327	0.338	0
UnitedKingdom	0	0	0.77
UnitedStates	0.136	0	0

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	68.095	68.345	66.704	66.679
Lagged outcome (1987)	68.445	69.243	68.221	66.829
Lagged outcome (1988)	69.318	69.972	69.176	67.203
Lagged outcome (1989)	70.652	69.846	68.560	68.895
Lagged outcome (1990)	71.711	71.180	70.682	70.238
Lagged outcome (1991)	72.120	71.292	70.978	70.137
Lagged outcome (1992)	72.009	71.608	71.323	72.161
Lagged outcome (1993)	71.915	71.521	72.215	72.580
Lagged outcome (1994)	71.197	71.091	71.393	73.261
Lagged outcome (1995)	71.319	71.973	73.068	73.384
Lagged outcome (1996)	71.609	71.872	73.026	74.470
Lagged outcome (1997)	72.243	71.866	73.184	73.988
Lagged outcome (1998)	72.374	72.282	73.227	74.962
Lagged outcome (1999)	71.816	73.243	74.415	76.385
GDP per capita	25.479	25.477	22.213	22.777
Share of children (%)	17.315	17.728	18.087	18.943
Share of the elderly (%)	13.419	15.843	15.726	15.664
Share of agriculture (%)	6.511	5.270	6.346	2.179
Share of industry (%)	33.692	29.198	28.755	29.087
Share of service (%)	59.796	65.533	64.799	68.734
GDP per capita (growth rate)	4.126	3.173	3.847	3.653
Population (growth rate)	0.326	0.367	0.328	0.264
Share of children (growth rate)	-2.637	-0.781	-1.044	-0.097
Share of the elderly (growth rate)	3.521	0.900	0.771	0.540

Table I.7: Outcome: female labor force participation rate (ages 50-54)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0	0	0
Belgium	0	0	0
Canada	0	0	0
Denmark	0.189	0.25	0.345
France	0	0	0
Italy	0.176	0.079	0.107
NewZealand	0	0	0
Norway	0	0	0.129
Portugal	0.191	0.163	0.199
Spain	0	0	0
Sweden	0.232	-	-
UnitedKingdom	0	0.461	-
UnitedStates	0.212	0.048	0.22

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	61.671	61.806	62.115	61.415
Lagged outcome (1987)	61.800	62.525	62.456	61.892
Lagged outcome (1988)	63.350	63.406	63.038	63.346
Lagged outcome (1989)	64.216	64.011	64.258	64.147
Lagged outcome (1990)	65.526	65.492	65.353	65.553
Lagged outcome (1991)	66.506	66.621	66.228	66.634
Lagged outcome (1992)	67.606	66.536	66.748	66.651
Lagged outcome (1993)	66.897	67.107	67.576	67.535
Lagged outcome (1994)	67.401	66.605	66.941	66.325
Lagged outcome (1995)	67.111	66.929	66.570	66.209
Lagged outcome (1996)	66.900	66.790	66.535	66.492
Lagged outcome (1997)	67.882	67.831	67.754	67.683
Lagged outcome (1998)	67.811	67.918	68.425	68.606
Lagged outcome (1999)	67.894	69.007	69.368	69.911
GDP per capita	25.479	24.429	22.430	24.880
Share of children (%)	17.315	18.642	18.672	18.848
Share of the elderly (%)	13.419	15.096	15.231	14.593
Share of agriculture (%)	6.511	6.663	5.489	6.994
Share of industry (%)	33.692	29.091	29.796	28.320
Share of service (%)	59.796	64.246	64.815	64.687
GDP per capita (growth rate)	4.126	3.580	3.817	3.491
Population (growth rate)	0.326	0.428	0.271	0.446
Share of children (growth rate)	-2.637	-0.838	-0.588	-0.758
Share of the elderly (growth rate)	3.521	0.861	0.651	0.676

Table I.8: Outcome: female labor force participation rate (ages 55-59)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0.16	0.473	-
Belgium	0	0	0
Canada	0	0	0.017
Denmark	0.047	0	0
Finland	0.15	0.198	0.291
France	0	0	0
Italy	0	0	0
NewZealand	0	0	0.12
Norway	0	0	0
Portugal	0.085	0	0.376
Spain	0	0	0
Sweden	0.094	0.329	0.196
UnitedKingdom	0	0	0
UnitedStates	0.464	-	-

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	49.863	50.187	50.489	50.848
Lagged outcome (1987)	50.806	50.640	51.522	50.685
Lagged outcome (1988)	50.919	51.550	52.017	52.105
Lagged outcome (1989)	52.208	52.932	52.897	53.244
Lagged outcome (1990)	53.944	54.066	54.152	54.459
Lagged outcome (1991)	55.500	54.725	54.979	55.372
Lagged outcome (1992)	55.556	55.269	54.811	55.271
Lagged outcome (1993)	56.404	55.283	55.097	54.865
Lagged outcome (1994)	56.359	56.602	55.988	55.646
Lagged outcome (1995)	56.965	57.053	56.945	56.623
Lagged outcome (1996)	58.088	57.869	58.617	58.233
Lagged outcome (1997)	58.670	57.949	57.486	57.383
Lagged outcome (1998)	59.070	58.635	58.121	57.892
Lagged outcome (1999)	58.667	59.958	59.529	60.042
GDP per capita	25.479	27.684	24.744	19.337
Share of children (%)	17.315	20.604	20.134	19.541
Share of the elderly (%)	13.419	13.281	13.950	14.310
Share of agriculture (%)	6.511	5.266	5.308	10.276
Share of industry (%)	33.692	26.479	25.959	29.771
Share of service (%)	59.796	68.222	68.634	59.953
GDP per capita (growth rate)	4.126	2.968	3.021	4.243
Population (growth rate)	0.326	0.858	0.830	0.389
Share of children (growth rate)	-2.637	-0.424	-0.437	-1.097
Share of the elderly (growth rate)	3.521	0.717	0.879	1.296

Table I.9: Outcome: difference in female LFP rate (ages 50-54 – ages 45-49)

Weights

Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0	0	0
Belgium	0	0	0
Canada	0	0	0
Denmark	0.189	0.25	0.345
France	0	0	0
Italy	0.176	0.079	0.107
NewZealand	0	0	0
Norway	0	0	0.129
Portugal	0.191	0.163	0.199
Spain	0	0	0
Sweden	0.232	-	-
UnitedKingdom	0	0.461	-
UnitedStates	0.212	0.048	0.22

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	61.671	61.806	62.115	61.415
Lagged outcome (1987)	61.800	62.525	62.456	61.892
Lagged outcome (1988)	63.350	63.406	63.038	63.346
Lagged outcome (1989)	64.216	64.011	64.258	64.147
Lagged outcome (1990)	65.526	65.492	65.353	65.553
Lagged outcome (1991)	66.506	66.621	66.228	66.634
Lagged outcome (1992)	67.606	66.536	66.748	66.651
Lagged outcome (1993)	66.897	67.107	67.576	67.535
Lagged outcome (1994)	67.401	66.605	66.941	66.325
Lagged outcome (1995)	67.111	66.929	66.570	66.209
Lagged outcome (1996)	66.900	66.790	66.535	66.492
Lagged outcome (1997)	67.882	67.831	67.754	67.683
Lagged outcome (1998)	67.811	67.918	68.425	68.606
Lagged outcome (1999)	67.894	69.007	69.368	69.911
GDP per capita	25.479	24.429	22.430	24.880
Share of children (%)	17.315	18.642	18.672	18.848
Share of the elderly (%)	13.419	15.096	15.231	14.593
Share of agriculture (%)	6.511	6.663	5.489	6.994
Share of industry (%)	33.692	29.091	29.796	28.320
Share of service (%)	59.796	64.246	64.815	64.687
GDP per capita (growth rate)	4.126	3.580	3.817	3.491
Population (growth rate)	0.326	0.428	0.271	0.446
Share of children (growth rate)	-2.637	-0.838	-0.588	-0.758
Share of the elderly (growth rate)	3.521	0.861	0.651	0.676

Table I.10: Outcome: difference in female LFP rate (ages 55-59 – ages 45-49)

Weights

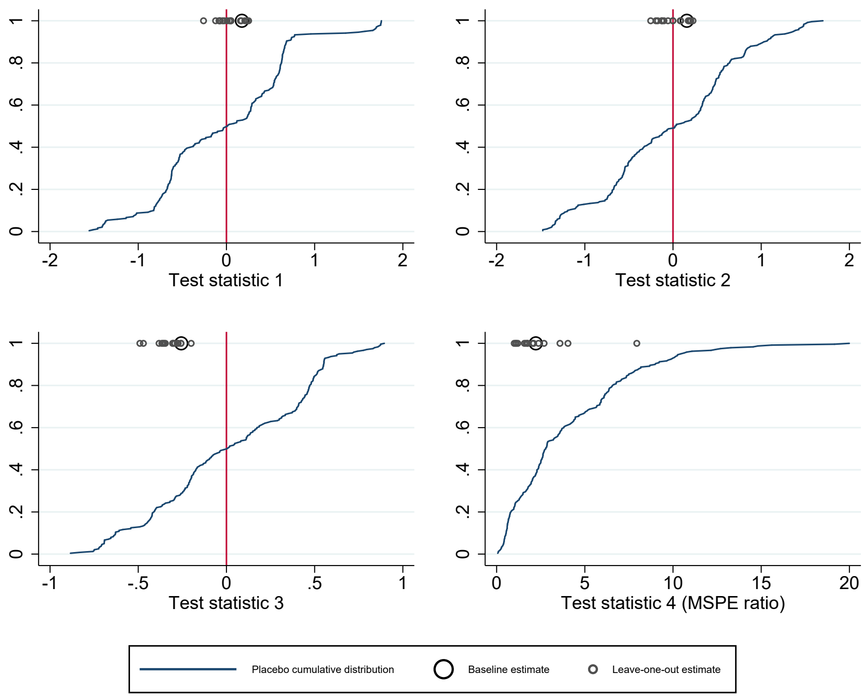
Country	Synthetic control 1	Synthetic control 2	Synthetic control 3
Australia	0.16	0.473	-
Belgium	0	0	0
Canada	0	0	0.017
Denmark	0.047	0	0
Finland	0.15	0.198	0.291
France	0	0	0
Italy	0	0	0
NewZealand	0	0	0.12
Norway	0	0	0
Portugal	0.085	0	0.376
Spain	0	0	0
Sweden	0.094	0.329	0.196
UnitedKingdom	0	0	0
UnitedStates	0.464	-	-

Predictor values

Predictor	Treated	Synthetic control 1	Synthetic control 2	Synthetic control 3
Lagged outcome (1986)	49.863	50.187	50.489	50.848
Lagged outcome (1987)	50.806	50.640	51.522	50.685
Lagged outcome (1988)	50.919	51.550	52.017	52.105
Lagged outcome (1989)	52.208	52.932	52.897	53.244
Lagged outcome (1990)	53.944	54.066	54.152	54.459
Lagged outcome (1991)	55.500	54.725	54.979	55.372
Lagged outcome (1992)	55.556	55.269	54.811	55.271
Lagged outcome (1993)	56.404	55.283	55.097	54.865
Lagged outcome (1994)	56.359	56.602	55.988	55.646
Lagged outcome (1995)	56.965	57.053	56.945	56.623
Lagged outcome (1996)	58.088	57.869	58.617	58.233
Lagged outcome (1997)	58.670	57.949	57.486	57.383
Lagged outcome (1998)	59.070	58.635	58.121	57.892
Lagged outcome (1999)	58.667	59.958	59.529	60.042
GDP per capita	25.479	27.684	24.744	19.337
Share of children (%)	17.315	20.604	20.134	19.541
Share of the elderly (%)	13.419	13.281	13.950	14.310
Share of agriculture (%)	6.511	5.266	5.308	10.276
Share of industry (%)	33.692	26.479	25.959	29.771
Share of service (%)	59.796	68.222	68.634	59.953
GDP per capita (growth rate)	4.126	2.968	3.021	4.243
Population (growth rate)	0.326	0.858	0.830	0.389
Share of children (growth rate)	-2.637	-0.424	-0.437	-1.097
Share of the elderly (growth rate)	3.521	0.717	0.879	1.296

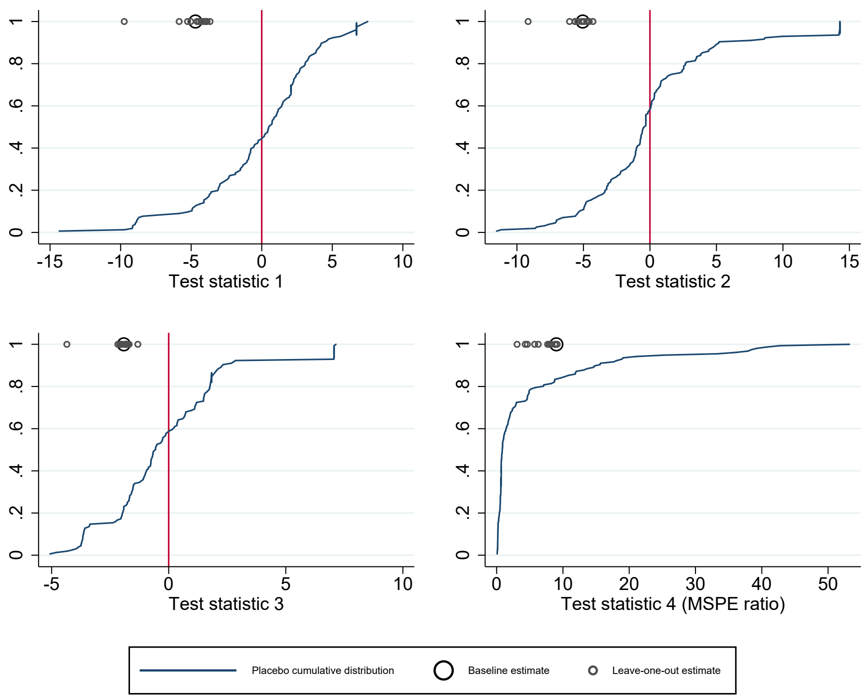
II Extended placebo tests for the other outcomes

Table II.1: Outcome: public health expenditure



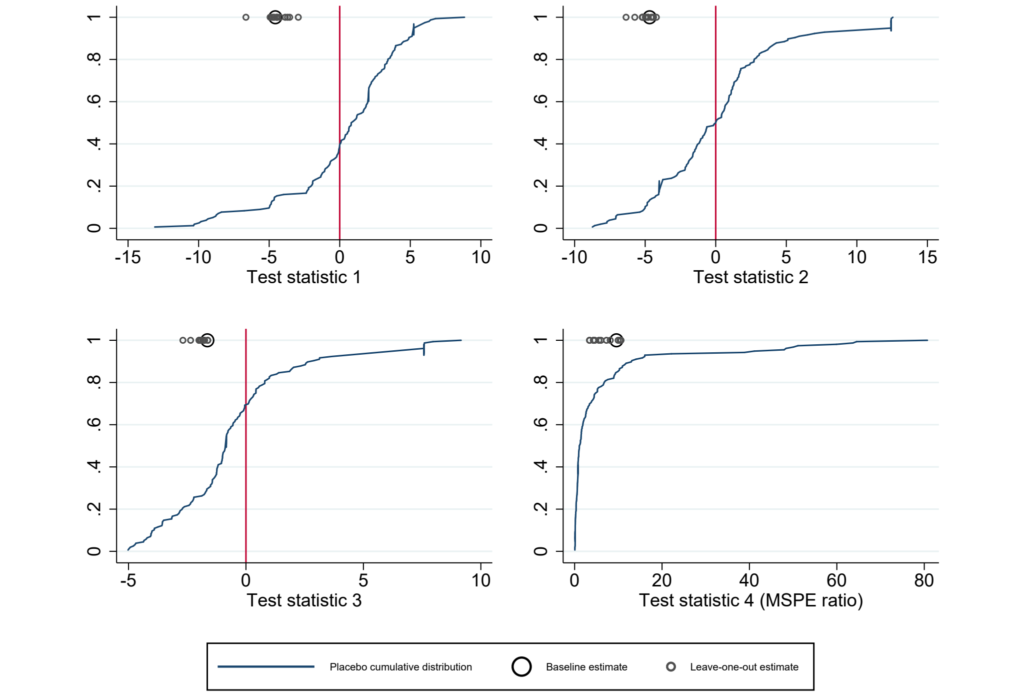
Note: see the note on Figure 6 for the estimation procedure.

Table II.2: Outcome: Female labor force participation rate (ages 40-44)



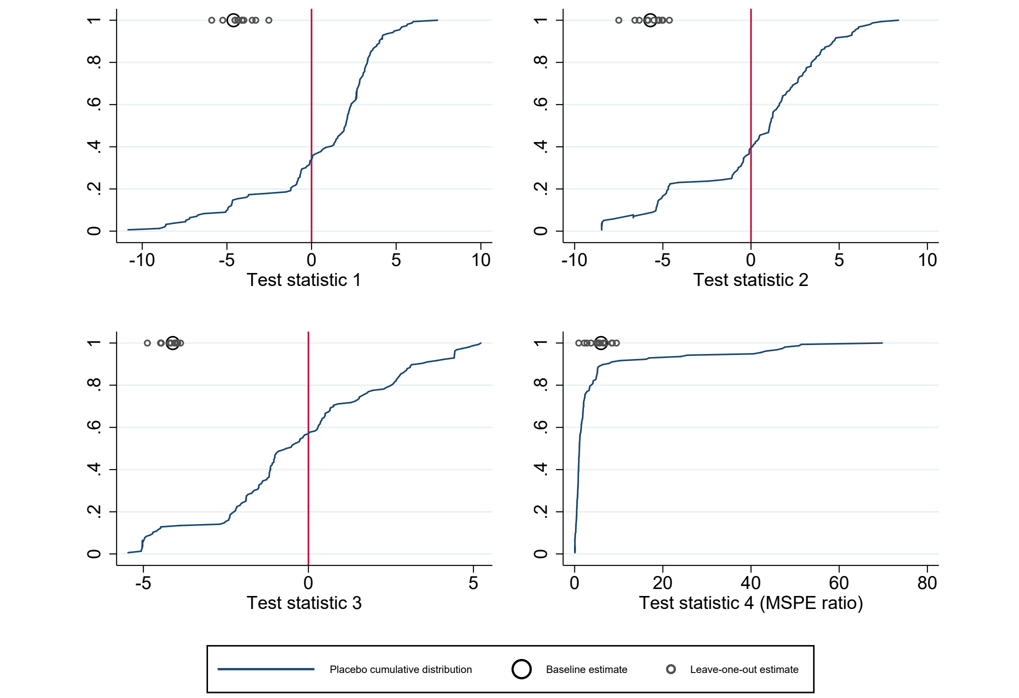
Note: see the note on Figure 6 for the estimation procedure.

Table II.3: Outcome: Female labor force participation rate (ages 45-49)



Note: see the note on Figure 6 for the estimation procedure.

Table II.4: Outcome: Female labor force participation rate (ages 50-54)



Note: see the note on Figure 6 for the estimation procedure.