UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THREE ESSAYS IN DEVELOPMENT & APPLIED MICRO

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HUIQIONG DUAN
Norman, Oklahoma
2018

THREE ESSAYS IN DEVELOPMENT & APPLIED MICRO

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ECONOMICS

BY

_____
Dr. Kevin Grier, Co-Chair

_____
Dr. Daniel Hicks, Co-Chair

_____
Dr. Scott Linn

_____
Dr. Benjamin Keen

_____
Dr. Jaeho Kim

_____
Dr. Joan Hicks

Acknowledgements

I would like to express my sincere gratitude to my committee members: Kevin Grier, Daniel Hicks, Benjamin Keen Joan Hicks, Jaeho Kim, and Scott Linn. This dissertation would not have been possible without their generous help. Special thanks to Kevin Grier and Daniel Hicks who have shown me how to become a qualified scholar. I would like to acknowledge my coauthor and co-chair, Dr. Daniel Hicks for his assistance and direction in Chapter 2 of this dissertation, *Education as Opportunity? The Causal Effect of Education on Labor Market Outcomes in Jordan*. Finally, I want to thank my parents Xinmin Duan and Taoxiang Wang, and my husband Weici Yuan. They have always been supportive through the whole process.

Table of Contents

List of Tables

List of Figures

Abstract

Chapter 1: China's National Poor Counties (NPCs) program represents one of the largest anti-poverty efforts ever undertaken in a developing country. In spite of this, no causal analysis has been undertaken to examine whether counties which graduate from the program actually achieve sustainable growth after support is removed. This paper provides the first long-term analysis of China's NPCs program on local development, investigating whether poor regions continue to develop economically after exiting the anti-poverty program. Focusing on the period 1986 to 2010, I construct plausible counterfactuals for counties that participated and graduated from the NPCs registry using the synthetic control method. This exercise reveals several new insights. First, on average, graduates of China's anti-poverty exhibit roughly the same growth outcomes as those still receiving support from the program. Second, I show that focusing on the average impact masks sizeable heterogeneity when disaggregating to a case by case analysis of counties. Importantly, for the subset of counties which graduated from the first wave of the poverty reduction program in 1993, there is a demonstrably faster expansion post-graduation. For more recent county's graduating from the program, however, I find significantly negative impacts of NPCs program completion in the long run.

Chapter 2: This paper studies the impact of the 1988/1989 educational reform in Jordan which extended mandatory schooling from nine to ten years and restructured secondary schooling. Despite weakness in the Jordanian labor market, our estimates suggest that an additional year of required schooling in the late 1980s was sufficient to improve labor force participation, employment, and wages. These effects were initially largest for women, while males with more education were also slightly more likely to be self-employed, work longer hours, and earn higher wages. We show that the extensive margin labor market gains we observe for men strengthen with age.

Chapter 3: This paper studies son preference among households in the United States. I present two analyses using the American Community Survey from 2012 to 2016. First, I estimate son preference in aggregate among parents in the U.S., contrasting immigrant populations to that of native born individuals. I consider evidence of two direct mechanisms – fertility stopping decisions and gender selection behavior. I show that there is no evidence of son preference among families in the U.S in aggregate through either mechanism, but there is indirect evidence of son preference -- mothers of daughters are less likely to be married than mothers of sons. Next, I analyze assimilation of immigrant gender preferences to U.S. norms relying on variation in duration of residence in the U.S., age of migration, and generation of migration. I find significant evidence to support the claim of son preference among first-generation immigrant populations originating from Asia and find that those son preferences were formed early in life, and once established, persist among first-generation immigrants. Higher order generations exhibit weaker gender

selection, consistent with assimilation to U.S. norms. Assimilation exhibits a high degree of heterogeneity across populations. Second and higher order generation Chinese immigrants conform to U.S. norms, while surprisingly, Indian immigrants exhibit even stronger son preference among higher-order generations.

Chapter 1. The Long Run Impact of China's Anti-Poverty Program: A Synthetic Control Analysis

1.1 Introduction

China has demonstrated remarkable success in poverty alleviation over the past three decades. The number of individuals living on less than $1 per day decreased by a staggering 442 million from 1984 to 2005. This reduction alone accounts for 96% of the decrease in poverty headcounts across the developing world over this period (Chen and Ravallion, 2010). Despite this progress, little is reliably known regarding the efficacy of government policies aimed at poverty reduction.[1] A clearer understanding of the mechanisms through which this progress has been made, is therefore relevant for continued development, for informing government resource allocation decisions, as well as for drawing generalizable lessons which may be useful for other developing counties (Ravallion, 2009).[2]

This paper seeks to address this gap by providing evidence regarding the long-term impacts of graduating from China's most ambitious anti-poverty campaign to date, the National Poor Counties (NPCs) program. Targeted at the county level and initiated by the Chinese central government in 1986, the NPCs program continues to provide substantial support to regions of the country today. Through the program, a subset of poor counties is targeted with government support for the purposes of reducing poverty. When a region is

[1] Loayza and Raddatz (2010) find that not only the rate of economic growth but also the growth structure matters for poverty reduction. Unskilled labor-intensive sectors like agriculture, construction, and manufacturing contribute the most to poverty alleviation. Rapid growth in agriculture significantly reduces poverty in Vietnam and rural areas of Indonesia (Janvry and Sadoulet, 2010; Suryahadi et al., 2009), however, the expansion in social security and social assistance transfers in Brazil are the largest contributors of the poverty reduction during 1984 to 2004 (Ferreira et al., 2010).

[2] There were 15% and 41% of people in South Asia and Sub-Saharan African living on less than $1.9 per day in 2013. Given the sheer scale of poverty reduction in the country, China's growth of income per capita is thus the primary determinant of global poverty reduction along this threshold. (Xavier, 2006).

considered developed, this support is withdrawn, and the county graduates from the program. Theoretically, over the long-term, a graduating county could demonstrate improved performance after graduation because it has been nudged into a stronger growth position, for example as individuals escape a subsistence poverty trap. Or a county could experience worse economic performance because regular governmental support has now been withdrawn.

Using the synthetic control approach of Abadie et al. (2003) and Abadie et al. (2010), I first demonstrate that on average counties which graduate from this program demonstrate long-term economic performance similar to non-graduates. Second, leveraging the synthetic control method to examine the impact on a period-by-period and county-by-county basis, I demonstrate that there is sizeable heterogeneity in the estimated impacts of graduating from NPCs designation across participants and over time. Specifically, the only robust evidence for long-term growth gains of the program appears to be concentrated among counties which graduated during the first wave of the poverty reduction program in 1993. For counties which graduate at later dates, I find that graduating from the program does appear to significantly hinder growth in these regions relative to their synthetic counterparts in the long run.

Several aspects of this analysis are novel. First, despite the massive scale of the NPCs designation program, few papers have studied China's poverty alleviation program, and those that do so have produced somewhat mixed results regarding its efficacy (Jalan and Ravallion, 1998; Park et al., 2002; Meng 2013; Park and Wang, 2010, Qin and Chong 2016).[3] This research reaffirms several findings in this nascent literature and builds on our understanding

[3] Park *et al.* (2002) and Meng (2013) find annual increase on income of 0.91% to 6.4% of the program within the disbursement period.

of the program's benefits and shortcomings. These results also add to analysis of other anti-poverty programs in china. For example, Ravallion and Chen (2005) find a sizable and statistically significant impact of World Bank's Southwest Program (SWP), a poverty alleviation program in China from 1995 to 2000, on mean income within the disbursement period (annually 9% increase on average).

Second, pre-existing research has focused on contemporaneous impacts of the program. This means that perhaps the most important question for assessing the success of a poverty reduction program – whether there are long-term impacts years after government support has been withdrawn, remains unanswered.[4] Indeed, the poverty reduction program aims to, not merely reduce the number of individuals living in poverty, but also to establish the sustainable development of NPCs through enhancing the local economy and improving infrastructure.[5] However, sustainable development implies that graduates of national poor counties registry continue growth after the external aid disbursement ends.

This paper fills this gap. Long-term impacts are critically important for understanding whether NPCs designation and participation in an anti-poverty program in general, can generate sustained economic growth after the program ends – i.e. in the absence of continued government support. To the extent that poor counties were stagnant prior to NPCs designation, these results are important for understanding whether the NPCs program can help counties escape subsistence poverty traps and transition toward sustained growth. Perhaps the closest analysis in this regard is Chen et al. (2009), which examines the medium-

[4] In many ways, this is conceptually the aim of anti-poverty programs in comparison to transfer schemes. Evidence of long-term impacts presented here thus helps answer this important policy question.

[5] It is emphasized by the central government in "Outline of China's Rural Poverty Alleviation and Development (2001-2010)" and "The National Poverty Alleviation Program (1994-2000)".

run impacts of the World Bank SWP anti-poverty program, finding sizable impacts which persist for four years after the external aid ended.

China has undertaken three waves of the NPCs designation poverty reduction program, wherein the government designated certain, qualifying counties as national poor counties. Once classified as such, the counties received generous financial support and benefits. For example, the central government provided a total of $14.9 billion (124 billion RMB) in the second wave of poverty alleviation program, annually accounting for approximately 5%-7% of total central government expenditures, and the total poverty funds, which mainly include subsidized loans, Food-for-Work, and budgetary grants, reached 232.7 billion RMB during 1986-2002 (Wang et al., 2004). The benefits of the program included subsidized loans, grants, tax reliefs, infrastructure construction and other investment privileges from central and local governments (Park et al., 2002). Those short-run programs may have long-term effects through the potential mechanisms, for example, improved infrastructure can support higher income growth in the future, the existence of health care system support healthier and more efficient labor, and lower tax burden provide higher disposable income for families to invest in small business. There were 258 national poor counties in 1986, while this number increased to 592 in 1994 and stayed constant thereafter.[6] Counties graduate from the program once they successfully pass one or more predetermined thresholds of poverty and new counties take their place immediately.[7]

Methodologically, this paper contributes more broadly to the literature evaluating regional anti-poverty campaigns, of which the NPCs program is an important example.

[6] The number of counties in China changed slightly from year to year. In 1994, there were 1735 counties; and in 2010, there were 1636 counties, as shown in Figure 2 and Figure 3.

[7] "Graduate" means that national poor counties left the anti-poverty program according to the fact that they passed the predetermined thresholds of poverty in 1993, 2000, or 2010. The decision of graduation is made by the central and provincial governments.

Specifically, I use the synthetic control approach to uncover the dimensions through which poverty reduction has occurred across space and time in China, leveraging this approach to answer multiple policy relevant questions.

First, through an aggregated synthetic counterfactual (Cavallo et al. 2013), I estimate the overall average treatment effect of NPCs graduating from the program. In comparison to counties designated as NPCs but which have not yet graduated, I find no robust difference in economic status 7 to 10 years later. On average, the treated group has only a 2.78% (2 to 85 yuan (RMB) based on 1500 to 3500 yuan (RMB)) annually higher net income per capita in graduated NPCs relative to the synthetic controls. Second, disaggregating by wave of the NPCs program, I present evidence of robust growth effects for the 1993 cohort of graduates, not evident for later cohorts. For example, seven years after graduation, the net income per capita of the treated counties in group 1, 2, and 3 increase by 91%, 42%, and 44% respectively since graduation, while the synthetic counterfactual sees a 44%, 53%, and 65% increase. These estimates contrast the impact of graduation for counties which participated in the NPCs with one another, alleviating concerns over selection into the program and the evolution of political favoritism on potential growth outcomes. If anything, political favoritism should bias towards findings negative impacts of the program, since graduating counties receive less support, and may thus be less of a focal point for the government.

Finally, using the synthetic control approach, I also rerun the analysis contrasting the post-graduation growth experience of NPCs against a second counterfactual built entirely from the subset of counties which never participated in the program. This counterfactual pool suffers less from issues of selection out of treatment and represents a desirable benchmark of economic activity in the economy, with the tradeoff of not being able to compare the impact of no-longer receiving aid to counties which do receive aid.

Relative to a synthetic of counties which never participated, the set of counties graduating in 1993 appears to demonstrate stronger economic performance 7 years out. Counties graduating at a later date exhibit slightly slower growth on average. As the results are similar when counterfactuals are constructed from the pool of non-participant counties, this provides additional confidence suggesting that the findings obtained when comparing NPCs which have graduated to those which have not yet graduated are not merely an artifact of unobserved factors which may drive selection into and out of the program such as government favoritism.

The remainder of the paper is organized as follows. Section 2 describes the three waves of poverty reduction program in China and describes research examining the contemporaneous effects of this program. Section 3 discusses the sample and data summary. Section 4 explains how the synthetic control method works. Section 5.1 presents estimates of the overall average impact of graduating from the NPCs program. Section 5.2 disaggregates these results by program wave, while Section 5.3 examines the results for individual counties. Section 5.4 explores the robustness of the results to the use of an alternative counterfactual pool. Section 6 concludes and discusses avenues for future investigation.

1.2 Anti-Poverty Policy in China: National Poor Counties

In the 1980s, China's economy was predominantly agrarian and suffered from high rates of rural poverty. In 1986, the central government formed the State Council Leading Group Office of Poverty Alleviation and Development (hereafter The Leading Group), an agency of the State Council and a specialized institution to direct and lead the anti-poverty policies nationwide. Through the Leading Group, the Chinese government initiated the first wave of poverty reduction program over the period from 1986 to 1993.

Initially, in 1986, the Leading Group designated 258 poor counties as national poor counties based on whether counties' rural net income per capita was less than 150 RMB in 1985.[8] These counties were targeted to receive support and transfers. By 1989, there were a total of 331 counties on the list of NPCs, and that list did not change until 1993.[9] From the maps of NPCs in figure 2, it is clear that there is a geographic pattern. NPCs in the first wave of poverty reduction program are more likely to be located in the middle part of China, while those that graduated (exited from the NPCs program) were more likely to be counties in the eastern region.

In 1994, the Leading Group adjusted the member of NPCs based on counties' rural net income per capita as of 1992.[10] As shown in figure 1, the total number of NPCs reached 592 and did not change during the entire period of the second wave of anti-poverty program from 1994 to 2000. As shown in figure 2, NPCs in the second wave of the program were concentrated in the middle and western regions, while counties which graduated from the program were predominantly located in the middle and eastern parts of the country. Graduation only occurred in the final year of each wave (1993, 2000 and 2010). There is no county's leaving that happens during each wave of poverty reduction program.

During the first two waves of alleviation poverty program, the central government provided substantial funds on an annual basis through three main channels: subsidized loans, Food-For-Work, and government budgetary grants.[11] Projects were expected to improve the

[8] The threshold for minority counties was 200 RMB in 1985 and 300 RMB for the revolutionary base counties.

[9] For political considerations, other 73 counties, mainly are revolutionary base areas and pastoral counties, were successively added onto the list of NPCs from 1986 to 1989.

[10] Counties with net income per capita below 400 RMB were designated as national poor counties, while existing national poor counties whose net income per capita exceeded 700 RMB (the threshold of subsistence) graduated from the program.

[11] Subsidized loans supported the poor households and enterprises of poor areas in industry and agriculture; Food-For-Work sought to develop infrastructure, such as constructing roads, building water storage facilities and afforestation to provide job opportunities to surplus farm labors; Government budgetary grants were

economy in NPCs and to directly assist poor residents. Nevertheless, these figures do not capture the full extent of the NPCs program as some less well documented funding also occurred outside these three main channels.[12]

The third wave started in 2001. During this wave, the Leading Group not only altered the set of counties designated as NPCs but also modified the criteria for selection. They abandoned the single poverty line criterion and used a new composite index to select NPCs, which consists of poverty headcount, rural net income per capita, local GDP per capita, and local fiscal revenue per capita.[13]

The contemporaneous impact of NPCs designation has been studied by a few researchers, each examining slightly different samples or time frames. Jalan and Ravallion (1998) find that in the first wave of the program, consumption growth is higher in areas covered by the program relative to uncovered areas. Park et al. (2002) show that the alleviation program had a significant, positive impact on rural income growth, with a 2.25% annual increase in growth rates from 1985 to 1992 and 0.91% from 1992 to 1995.

Meng (2013) applies a fuzzy Regression Discontinuity (RD) Design method to evaluate the overall effectiveness of the second wave of poverty alleviation program during the period of 1994-2000. He finds that the overall effect of the program was to promote rural net income per capita by roughly 38% or 6.4% per year. Park and Wang (2010) estimate the impact of poverty alleviation program using matching methods on panel data from 2001 to 2004, studying household and village level income and consumption. They find that the

mainly used to improve primary education and basic health care, knowledge and technology diffusive, promotion of culture and training in poor areas.

[12] Total poverty funds include all sources of anti-poverty funds like funds from provincial governments and international poverty funds, but not just the three main channels. The percentage of poverty funds through the three main types of anti-poverty projects was about 96% of the total poverty funds for the first wave of the program.

[13] More details regarding China's anti-poverty program are available upon request.

program did not affect household income and consumption for the poorest but increased richer households' income and consumption by 6.1% to 9.2%. However, Qin and Chong (2016) find the ineffectiveness of the third wave of national poor counties program. The relative growth rate of GDP, local government revenue and income of participant counties were found to be negative in comparison to non-participant counties. Seems like, previous literature find positive effect for the first two waves of the program, while negative effect for the third wave of the program.

Since this article estimates the long-term impacts of China's poverty reduction program, it serves to complete Park et al. (2002) and Meng (2013) but also to generate estimates of unique policy relevance and importance. Observing some contemporaneous impacts is unsurprising given that the poor counties receive generous financial aid packages, as we should mechanically expect to witness increases in short-term consumption and income from the transfers.

As such, it seems inadequate to judge the merit of China's poverty alleviation program based solely on the short-run effects; because if the graduating counties stagnate or regress in terms of growth and poverty reduction, then it suggests the NPCs program is nothing more than a transfer or subsidy. If instead, they realize sustain growth, then it implies that the program may have fundamentally altered the counties, achieving lasting poverty reduction. In this subsequent section, I emphasize sustainable growth after graduation, a fundamental goal of any poverty reduction program.

1.3 Sample and Data Summary

In this paper, I consider the effects of poverty reduction policies undertaken at the county level for consistency, even though a few administrative county-level cities and

suburban districts of the major cities were designated as NPCs.[14] Among NPCs, three joined the program twice. They were designated as NPCs during the first wave of the program, then graduated from the program as non-poverty counties in 1993, and were re-assigned to the program in 2001.[15] Anecdotally, the existence of these counties serves to demonstrate the motivation for this article's analysis: namely the question of whether China's poverty alleviation program establishes sustainable growth is an important one as graduating counties may regain high levels of poverty in the long-term.

Moreover, the synthetic control method provides the specific effect for each treated unit instead of only the average effect for the treated group, which provides the access to identification of heterogeneous effects. This particular strategy is thus well suited to measure the long-term effect of a treatment, which requires that I track the evolution of the outcome variable for many years post-treatment. Therefore, the feasible sample is limited to poor counties that graduated from the first two waves.[16]

There are several technical nuances worth noting when studying poverty alleviation. Duclos et al. (2010) argue that any analysis of poverty would be biased if the estimators of interest are non-linear across time periods. Therefore, in addition to looking at any aggregate impacts, I also divide the sample into three sub-groups, according to the different entry time and durations of poor counties in the program, to estimate the overall average effects for

[14] Counties whose designation changed to be county level cities or suburban districts of big cities in the duration of aiding poor policies are not included in my sample, because the change suggests those regions are different from other NPCs at the administrative level. These changes were not uncommon with 21 cases in the first wave of poverty reduction program; 32 cases in the second wave, and 24 cases in the third wave. In China, cities are usually treated as the upper administrative division of counties, but a few regions are designated as county-level cities. To the extent that prosperous regions become upgraded to city level regions, the results I observe may understate the true impact of the NPCs program.

[15] They are Guidongxian, Ruchengxian, and Chengbumiaozuzizhixian. Based on the available data, from 1994 to 1998, the average growth rate of net income per capita for Guidongxian and Ruchengxian is 22.5%, while the average growth rate for non-participant counties and NPCs are 63% and 45%.

[16] The third wave of poverty reduction program was ended in 2010.

each group and also each wave. As shown in Table 1, they are here define as group 1, counties that entered the program in 1986 and graduated in 1993 with 8 years' benefits from the program, which means they participated in and then graduated from the first wave of the program; group 2, counties that were designated as NPCs in 1994 and graduated from the program in 2000, or they participated in and graduated from the second wave of the program; and, group 3, counties that stayed in the program from 1986 through 2000, benefiting for 15 years, or it represents counties that participated in the first wave of the program and graduated from the second wave.[17]

Since I want to test whether the growth trend of NPCs changes with the end of external aid disbursement, I have two potential control groups for the choice of counterfactuals. One is the NPCs that participated in the program at the same time of the treated counties entry but remained in the program as the treated counties graduated. Another one is the non-participant counties that never have been designated as NPCs in China. Even though the synthetic control results are similar for the two potential control groups, the former one is more appropriate according to the criterion of donor pool of synthetic control, which restricts the control group to counties that have similar characteristics as the treated counties in pre-treatment period.[18]

I obtain county level data from China Statistical Yearbooks Database of CNKI (China National Knowledge Infrastructure). The database was generated by Tsinghua University and Tsinghua Tongfang Company in June 1999 and covers almost all published

[17] Previous literature on the effects of the second and third wave of poverty reduction program does not distinguish the effects of the previous waves of poverty reduction program from the estimated effects of the latter two waves, given that NPCs in the second and third wave may also have taken part in the previous waves of the program.

[18] The results of synthetic control with non-participant counties as control group are included in section 5 as robustness check.

Yearbooks in China at the national, provincial, city, county, and industry levels. Further statistics are derived from official Yearbooks like China County (City) Social and Economic Statistical Yearbook (1996, 2000-2011), China Statistical Yearbook for Regional Economy (2000-2011), China County Statistical Yearbook (1980-1991), China rural statistical yearbook (1987-2011), China's Ethnic Statistical Yearbook (1987-2011) as well as provincial yearbooks, cities' yearbooks, and counties' yearbooks.

The main criterion for designating as a national poor county is rural net income per capita, which measures the average net income level of permanent residents within a given rural area. Net income is defined as gross income (including cash and income in kind) less taxes, fees, operating expenses of family businesses and depreciation of fixed assets. Data on rural residents' income is derived from national surveys conducted by the National Bureau of Statistics.

Table 2 presents summary statistics for non-participant counties, comparing NPCs to all others from 1986 to 2010. As expected, the results in Table 2 suggest that NPCs' average GDP per capita, population, area per capita, fiscal revenue per capita, residents' savings balance per capita, and rural net income per capita are significantly lower than the non-participant counties' averages. NPCs are more likely to be bordered, pastoral, minority, and mountainous counties. On comparing the final year within disbursement period with the first year, the growth in savings rate is much higher for NPCs relative to the national average. The high savings rate finding is consistent with Chen et al. (2009), which implies that the sizable increase in income during the disbursement period is transient (it was saved) when we interpret it in Permanent-Income Hypothesis.[19] Table 2 also provides similar summary

[19] The uncertainty of the impact of poverty reduction program in national poor counties might lead to precautionary saving.

statistics for graduated NPCs and ungraduated NPCs from 1986 to 2010. Table 2 shows that graduated NPCs have higher GDP per capita, population, fiscal revenue per capita, residents' savings balance per capita and net income per capita than ungraduated NPCs on average. This table also shows that graduated NPCs have relatively lower average area per capita, the ratio of primary industry added value/GDP, and less likely to be a border, mountainous and minority counties.

The average rural net income per capita for NPCs in 1986 was 786.41 RMB, about 68% of the national average. In 2010, the average net income for NPCs was 2,710, more than tripling the value in 1986, a statistic, which, when viewed in isolation, gives the impression that the poverty reduction program must be a success. However, a closer look reveals that the country average grew at a comparable or, in some cases, faster rate. Figure 3 depicts the trends of rural net income per capita for both the NPCs and non-participant counties. It shows similar trends, yet the line of NPCs always lies below the non-participant counties line, which implies that the perceived success of the poverty relief may merely be a byproduct of China's fast nationwide growth (Montalvo and Ravallion, 2010).[20] The NPCs line converges to the non-participant line before 1994 but diverges after 1994, which supports the argument that the impacts of poverty reduction program on income growth may have declined in the 1990s in Park et al. (2002).

Figure 4 plots the gap of average net income per capita between graduated NPCs and non-participant counties for each group. Figure 4.A depicts that the gap between graduated NPCs in group 1 and non-participant counties is narrower over time, or that the trend of average net income per capita for graduated NPCs converges to the trend of non-

[20] Higher economic growth periods were associated with higher rates of poverty reduction periods during 1981 to 2001 in China (Ravallion and Chen, 2007).

participant counties. In 1999, the graduated NPCs had a higher average income level than non-participant counties. It means that the poor counties that graduated in group 1 performed as well as or even better than other non-participant counties on average after they graduated from the first wave of poverty reduction program in 6 years. Figure 4.B shows the common trend of average net income per capita between graduated NPCs in group 2 and non-participant counties from 1994 to 2000, but the gap between them enlarges from 2001 to 2010. It means that, on average, the graduated NPCs in group 2 had similar income growth as non-participant counties when they were receiving the aid disbursement, but were not as good as non-participant counties when the external aid disbursement ended. From Figure 4.C, I find the same results as in Figure 4.B for graduated NPCs in group 3. The two trends parallel with each other from 1986 to 2000 but diverge from 2000 to 2010. Prima facie, graduating counties appear to have performed worse relative to non-participant counties after aid flows are suspended. Section 4 explores this relationship causally.

1.4 Empirical Strategy

1.4.1 Treatment

I study the long-term impact of the NPCs program on income. In particular, I am interested in whether poor counties sustain growth, after graduation, at a pace comparable to that achieved during the disbursement period. If the impacts of anti-poverty program die out just after the poor counties' graduation, one can regard the NPCs program as just providing immediate extra funds for those poor counties within the disbursement period, but may have little impact on poor counties' economic structure, labor structure, and local infrastructure, having little impact on poor counties' growth capacity. One can conceptualize this possibility as the poverty alleviation program merely giving a man a fish,

rather than teaching a man to fish. If instead the counties grow sustainably in comparison to those still receiving aid, such evidence suggests that the program may have helped the region change structurally, for example to escape poverty traps.

The treatment effect in this article is the long-term impact of participating in the anti-poverty program after having received external governmental aid for seven to fifteen years. Therefore, the key event or intervention in interest is the graduation of poor counties from the NPCs program. That is to say, the primary focus of the analysis is NPCs regions that participated in and then graduated from the poverty reduction program (for simplicity, these can be thought of as treated counties).[21] Establishing the treatment effect in such a manner facilitates an analysis of the program's effect on sustained, long-term growth. The relationship among counties, NPCs, graduated and ungraduated NPCs, and non-participant counties are reported in Table 3. Table 3 also illustrates the particular treatment group and control group for the short-run and long-run impacts estimation.

In 1993, 20 of the counties which had entered the NPCs registry in 1986 graduated from the program (group 1). In 2000, 74 poor counties exited the program, but not all "graduated."[22] A compulsory policy forced out 33 poor counties from five wealthy, eastern provinces (Liaoning, Shandong, Zhejiang, Fujian, and Guangdong).[23] Due to the inherent uncertainty regarding valid graduation, I omit the aforementioned counties from the analysis. This leaves as the treated sample, 41 graduating counties in 2000, of which 27 originally started in 1993 (group 2) and 14 had participated in the program for its full existence – the

[21] Unlike the existing literature, which views entry into the national poor counties registry as treatment, in this article, counties that participate in and then graduate from the registry compose the treatment group.
[22] The graduation rate, 20 out of 331 and 41 out of 592, were low for both waves of poverty reduction program.
[23] These particular counties exited based solely on the compulsory political policy and the fact that they are in those five wealthy provinces, but not according to the graduation criteria.

15 years from 1986 to 2000 (group 3). To summarize, these three sets of treated counties include 20 poor counties in group 1, 27 in group 2, and 14 in group 3.

1.4.2 Empirical Strategy

In this article, I use the synthetic control approach developed in Abadie et al. (2003) and formalized in Abadie et al. (2010) to evaluate the long-term causal effect of poverty reduction program. The synthetic control method has been applied to many different topics, such as development, politics, and labor (Cavallo et al., 2013; Abadie et al., 2014; Dube and Zipperer, 2015; Grier and Maynard, 2016). Below I sketch an outline of the synthetic control method (SCM) in this context. A more technical description of the methodology can be found in Abadie et al. (2010).

Identification of the causal effect of a policy is often difficult to ascertain because of selection effects- for example because central planners do not assign policy randomly. In this analysis, counties that graduated from poverty reduction program likely differ fundamentally from those that remain (or from those that never participated). A simple direct comparison between treated and untreated counties will generate biased estimates.

Ideally, the problem could be overcome if I could simultaneously observe the graduate counties both with and without the treatment of intervention.[24] Since this is impossible, one alternative is to construct a counterfactual for the treated counties to answer the following question: how would we expect the treated counties to fare, had they not left the poverty alleviation program? One way to create a counterfactual is through the SCM, which is based on the notion that a combination of comparative counties generally resembles

[24] Suppose we have a population of i=1, 2,…, N and Di=1 (Treatment) or Di=0 (Control). Then, the treatment effect relative to control for county i is, $T_i = Y_{1i} - Y_{0i}$, where Y_{1i} is Y_i if Di=1 and Y_{0i} is Y_i if Di=0; T_i is the causal effect.

the characteristics of a treated county better than any single comparison county alone. This methodology is particularly suited for comparative case studies given that it analyzes only one treated county at a time.

If the constructed counterfactual closely matches the unit of interest prior to the treatment event, in theory, a synthetic county would be expected to have the same behavior as the treated county would have had in the post-treatment period if it were to keep receiving external aid. Thus, I can estimate the causal impact by contrasting the post-treatment difference between the treated (graduating counties) and their synthetic controls.[25]

This method in many ways resembles the difference-in-differences approach with the assumption of common trend being essential for plausible inference. Their major differences are that, first, difference-in-differences only requires the treated and the control groups to have parallel pre-treatment paths of the outcome variable (i.e. common trend) while synthetic control attempts to make their paths overlap with each other. Second, difference-in-differences analysis typically produces an aggregated average of every unit in the control pool while the synthetic control method utilizes a weighted average of the most relevant matches and as such only a few units are typically assigned positive weights. Finally, unlike difference-in-differences, accounting for only the time-invariant unobservable missing variable bias, synthetic control can account for the time-variant unobservable omitted variable bias, which is important to the identification of causal effects of China's anti-poverty program given that the anti-poverty program occurred in three waves and the treated counties receive treatment at different periods.

A key criterion of the selection of comparative counties for the donor pool is that the control counties should have similar characteristics as the treated counties in pre-

[25] Technical proof and details can be found in Abadie et al. (2010).

treatment period and those control counties should not be affected by the treatment in post-treatment period.[26] Thus, I first select a donor pool of NPCs that entered the program at the same time as graduated NPCs, but did not leave the program when the treated counties graduated. Moreover, each national poor county received similar amount of anti-poverty funds per year on average, which supports that the remaining NPCs are good comparative units for the treated counties. As an additional exercise, I consider an alternative pool of donor counties comprised of "non-participant counties" – non-participant in the sense that they never participated in the NPCs program.[27] This provides a counterfactual of economic growth in China among similar regions in the post-treatment period, and can thus also provide useful inferences about the post-graduation path of participants. These results can be found in Section 5.4.

The SCM constructs a synthetic counterfactual by creating a weighted average of all potential comparison counties, which reproduces the pre-treatment characteristics of the treated unit, including the trajectory of the outcome variable (rural net income per capita in this article) and the average values of the variables that have predictive power for the outcome variable. To optimize the weights such that the synthetic control most closely approximates the treated county, Abadie et al. (2010) recommend an algorithm to minimize the mean squared prediction error between the treated unit's outcome variable as well as predictors and the synthetic's in pre-treatment period. Furthermore, in computing weights for the untreated units, one should ensure that the treated unit's data falls within the convex

[26] To avoid interpolation biases, Abadie et al. (2015) require the care restricting the units of donor pool with characteristics similar to the treated units.

[27] I am grateful to two anonymous referees for the suggestion of this potential donor pool as an additional way of conceptualizing the long-term impacts of graduation from the NPCs program.

hull of the control units' data. That is, all weights are non-negative and sum to one.[28] This exercise guards against extrapolation bias.

Recall that in this setting, treated counties graduate from the program, while control counties remain NPCs designated and continue to receive support. Since categorization is based on the economic performance of the county, treated counties, by definition, outperform all others and thus may not fall within the convex hull of the data. This could jeopardize the validity of the exercise empirically; however, several factors suggest that the SCM is likely still appropriate in this context.

First, the criterion of graduation from the poor reduction program in 1993 depended only on the level of net income per capita in 1992, not on the growth trajectory of income from 1986 to 1993.[29] Second, the central government did not strictly follow the criterion of graduation in 1993 to decide which counties should leave and which counties should stay in the second wave of the program.[30] Park et al. (2002) find that political factors like lobbying efforts influenced the designation of NPCs. Moreover, Meng (2013) states that administrators may rely on other unobservable variables to designate the second wave of NPCs. Third, the central government provides different criterion for designating as NPCs

[28] Through synthetic control, I construct a counterfactual for each i when Di=1, which is $\left(\hat{Y}_{0i}\middle|D_i=1\right)$, such that $\sum_{j=1}^{J} w_j^* Y_{0jt_0} = Y_{1it_0}$ and $\sum_{j=1}^{J} w_j^* Z_{0jt_0} = Z_{1it_0}$ where w_j^* is an element of the vector of weights $W = \left(w_1, \ldots, w_J\right)'$ such that $w_j \geq 0$ and $w_1 + \cdots + w_J = 1$; j=1, …, J represent the units in donor pool; Y_{0jt_0} is the outcome of control unit j in pretreatment period and Y_{1it_0} is the outcome of treated unit i in pretreatment period; t_0 represents the pretreatment period; Z is a vector of observed covariates.

[29] The criterion for graduation from poverty reduction program in 1993 is that the net income per capita of poor county in 1992 was larger than 700 yuan (current price).

[30] The net income per capita of graduated counties like Gaolanxian and Zhengningxian in Gansu province, Linquxian and Yiyuanxian in Shandong province were less than 700 yuan in 1992. However, the net income per capita of staying counties like Lianchengxian, Shanghangxian, Wupingxian, and Changtingxian in Fujian province, Guangchangxian in Jiangxi province, Balinzuoqi, Wushenqi, Balinyouqi, and Keshiketengqi in Neimenggu province, Yanchixian in Ningxia province, and Yinanxian in Shandong province were larger than 700 yuan in 1992.

for minority counties, "revolutionary base" counties, and other normal counties.[31] The arbitrary nature of these assignments suggests that economic considerations are not the sole determinants. Fourth, the relative income level ranks of NPCs change significantly over time. I use a transition matrix to estimate the frequency of the transition. Table 4 reports the mobility of relative income ranks of poor counties from 1986 to 1992. I find that high income per capita counties in 1986 do not necessarily remain in the highest net income per capita group in 1992. Furthermore, low income per capita counties in 1986 could become a part of the highest income level group in 1992. This jostling in positions suggests that creating a synthetic control for the counties that graduated by using the counties that did not graduate as donors should be a viable exercise. Finally, in practice, I empirically find a set of control counties for each treated county which produce a close match of pre-treatment observables.

For the second wave of poverty reduction program, each province government determines the graduation criterion, and the central government only limits the number of NPCs that each province has. Although this means that finding valid donors within a province is less likely, there still exists a sufficient across-province variation to obtain suitable donors to legitimately employ synthetic control.

There is no conventional method for inference (e.g., z-tests or t-tests) in the SCM. Instead, I carry out a placebo test proposed in Abadie et al. (2010) to compute the significance level and to enhance the credibility of the findings. This test, in essence, is a type of permutation test, which is nonparametric and does not entail distributional assumptions about the data. Functionally, I test whether or not the growth effect observed could also be purely due to randomness. That is, there could be some random variation in the growth of

[31] The poverty line for non-participant counties in 1985 was 150 yuan, which was the criterion of designation in 1986. However, for political consideration, the cutoff for minority countries was 200 yuan and for "revolutionary base" counties were 300 yuan.

rural net income per capita not associated with the impact of poverty reduction program across counties. Perhaps the observed post-treatment gap between the treated and the synthetic county exists simply because the graduate county happens to be the one that grows faster in post-treatment, a phenomenon which might occur absent treatment. Thus, I estimate the probability of witnessing such a gap when the treatment does not occur simply by shuffling the labels of treatment and controls, and recalculating the gap on each permutation.

More specifically, I use the same synthetic control procedure (the same pre-treatment period and the same list of indicator variables) to form a synthetic counterpart for every county in the donor pool and record their gaps. I then put all gaps together in a graph and rank them. For example, assume there are 49 units in the donor pool, and if the estimated synthetic effect is larger than any of the 49 placebo effects, this is a strong indicator that treatment results in an outsized impact in the post treatment period, and one could nominally assign a pseudo significance level based on the rank of 1/50=2%.

To estimate the average long-term impacts of poverty reduction program for each group, I compute the post-treatment average effect and the statistical significance of estimated effects for each group by following the method of Cavallo et al. (2013). This new technique allows my analysis to contrast the average impact with that for individual groups and counties. For each group, I calculate the estimated average effect for each post-treatment year and follow the four steps which are introduced in Cavallo et al. (2013) to construct a distribution of average placebo effects; and then I can find the significance level for the estimated average effects. To increase the comparability of the placebo counties, I restrict the donor pool, which is used to compute the significance level of the estimated average effects, by match quality. Only controls that match as well as treated counties in pre-

treatment are included in the placebo pool to estimate the distribution of average effect. (Cavallo et al., 2013 and Galiani and Quistorff, 2016)[32]

1.5 Results

For counties leaving the program in 1993 and 2000, I examine the growth of net income per capita one to seven and ten years after their graduation. These time windows are potentially wide enough for the program to take effect and long enough to capture the long-term effects after graduation. The number of pre-treatment years considered depends on the duration of poverty reduction program and data availability. Theoretically, the longer the trajectory of the outcome variable is matched, the better the synthetic counterfactual may mimic the treated county. I attempt to fit 15 pre-graduation years of rural net income per capita when possible for group 3 counties, fit 7 pre-treatment years of net income path for group 2 counties, which participated in the aiding poor program in 1994 and graduated in 2000, and fit 8 pre-graduation years for those that joined the program in 1986 and left in 1993. I eliminate cases whose net income observations are less than five years before graduation.

To construct the synthetic control, I include available observables such as net income, the variable of interest, but supplement these criteria with net income predictors. I describe the predictor variables in greater detail in the following paragraph. Since I do not have a complete set of data on rural net income per capita and net income predictors for all counties, the final number of feasible cases is 7 out of 20 for the first treatment group, 20 out of 27 for the second treatment group, and 8 out of 14 for the third treatment group.[33]

---

[32] In individual synthetic control case study, I use the full donor pool to estimate the pre-treatment root mean squared errors (RMSE).

[33] I run synthetic control for counties that I can analyze. I will discuss how the number of cases in my sample affects the final results in the following section. If I reduce the number of years of outcome variable or

The pre-intervention characteristics include outcome variable for 4 years in pre-treatment period (the first and the last year, and other two middle years) and outcome variable's determinants, average values in pre-treatment period, which includes GDP per capita, the ratio of primary industry added value over GDP, fiscal revenue per capita, population growth rate, CPI, dummy variables of county types: hilly county, border county, pastoral county, minority county and mountainous county. [34],[35] I use the aforementioned variables to construct the synthetic counties in each group.

1.5.1 Average Impact of Graduating from National Poor Counties Designation

To better understanding and comparing the results of long-term impacts of NPCs program with the short-term results, I report the overall average results first though they are estimated based on the individual synthetic control results. Figure 5 shows that the synthetic counterfactual closely matches the treated group in pre-treatment period, and the values of net income per capita of treated group are slightly higher than the values of control group in seven years after graduation. On average, the treated group has 2 to 85 yuan (RMB) higher net income per capita than the synthetic counterfactuals in seven years after graduation, and that is a 2.78% annually higher net income per capita in graduated NPCs relative to the

predictors to have more cases, the synthetic counterfactual would not match to the actual outcome; and if the synthetic county fails to fit the trajectory of rural net income per capita prior to graduation, then the gap in post-graduation net incomes between the real and synthetic counties could simply result from lack of fitness and therefore cannot be fully attributed to the lasting effect of the aiding poor program.

[34] The reason for only including 4 years but not all years in pretreatment period is to make sure that other predictor variables would receive large enough weights in the weighting matrix V of indicator variables; otherwise, using all years renders other predictor variables unimportant (Kaul et al, 2015). Those four years are 1986, 1989, 1991 and 1993 for cases in group 1; 1994, 1996, 1998 and 2000 for cases in group 2; 1986, 1991, 1996 and 2000 for cases in group 3.

[35] The reason for including the ratio of primary industry added value over GDP as part of predictors is that growth in primary sector is more important to reduce poverty than other sectors (Ravallion and Chen, 2007; De Janvry and Sadoulet, 2010). The reason for including fiscal revenue per capita is that it is part of the criteria of designation and graduation of national poor counties in wave 3. Moreover, the revenue-generating ability of local governments is a measurement of whether local governments can provide sustainable supporting funds for lasting growth after external aids end.

synthetic controls. While, from eight to ten years after graduation, the synthetic counterfactual surpasses the treated group, which is due to the fact that I only estimate seven years post-treatment effect for group 1 counties which have significantly positive long-term effects.[36]

1.5.2 Average Impact of Graduation by Class of Participant

I find heterogeneity in the results across groups, for instance, positive impacts are dominated in group 1, but group 2 and group 3 are dominated by significantly negative effects. Figure 6.A, 7.A, and 8.A graph the behaviors of the average graduate counties for each group compared to their synthetic counterfactuals. The synthetics closely mimic the averages in the pre-treatment period, which provides evidence that the constructed synthetics are valid counterfactuals for averages. Following treatment, figure 6.A suggests that the program created lasting impacts upon net income per capita for graduated NPCs from wave 1; the actual average surpasses the synthetic, a difference that grows over time.[37] Seven years after graduation, the net income per capita of the graduate counties is 91% higher than it was at the time of graduation, even though the synthetic predicts a 44% increase, had the counties not graduated. Figure 6.B presents the significance levels for the results in group 1. Therefore, the results indicate that the first wave of poverty reduction program, on average, caused a statistically significant increase on net income per capita in all seven years following graduation, with the exception of the 1st and 2nd years.[38]

[36] I calculate the overall average effects based on the group average. Graduated NPCs in group 3 experience, on average, significantly negative long-term effects seven to ten years after graduation.
[37] The actual averages are higher than the synthetic averages seven years after graduation except the 1st year.
[38] The impacts are significant at 10% for all the seven years after those poor counties graduated except the first and the second year.

In figures 7.A, I observe the synthetic outperforms the actual wave 2 average for the first nine years post-graduation, suggesting this wave may have been less effectiveness in creating sustainable growth. However, following this period, the actual wave 2 average converges to the synthetic but insignificant. For instance, seven years after graduation, the net income per capita of the treated counties increases by 42% since graduation, while the synthetic counterfactual sees a 53% increase. The negative impact, lower net income per capita, of the first, fourth, fifth and sixth year are significant at 10% as can be seen in figure 7.B. Again, figure 8.A depicts that the average net income per capita for graduate poor counties in group 3 are higher than it in synthetic counties after they graduated in the first three years and are lower after they graduated for more than three years. Seven years after graduation, the net income per capita of the graduate counties is 44% higher than that at the time of graduation, although the counterfactual scenario experiences a 65% increase. Those impacts are significant at 10% for all years after graduation except the 4th and 5th years. Thus, as discussed in above, only counties that graduated from wave 1 significantly outperformed, on average, the constructed counterfactual. While counties that graduated in 2000 performed well or at least the same as the synthetic controls only last for three years after graduation.[39]

Taken at face value, the above results suggest that only the first wave of poverty reduction program, on average, has a positive, lasting impact on net income per capita for graduate counties. In addition, the above results are consistent with the results of the gap between graduate counties and non-participant counties as well as the results in Chen et al. (2009).

[39] The long-run effect of the second wave of the program is consistent with the results in Chen et al. (2009); the long-term impacts of World Bank's Southwest Program only last for four years after external aids ended.

I offer four possible explanations for the results. First, the criteria for poor counties to graduate were more clear and more objective for the first wave of poverty reduction program while the criteria were more nebulous and more subjective for the second wave of poverty reduction program, which led to the positive, clear average results for counties in group 1 and ambiguous results for counties in group 2 and 3.[40] Second, political factors like lobbying efforts of counties' government affect the process of designating and graduating NPCs. Although designation as a national poor county is not a positive signal for a county to attract outside investment, it can receive substantial financial support from central and provincial governments. Counties, including qualified NPCs and unqualified counties, have the incentive to lobby and influence the decision of NPCs. So, it is politically difficult to remove NPCs from the program registry, given that the leaders of those counties have strong incentive to fight to retain the access to the generous financial support from governments. This may explain why I find some poor counties that had passed the predetermined thresholds of poverty still stayed in the program, while some poor counties that had not passed the thresholds graduated from the program. Combine the nebulous and subjective criteria of graduation in 2000 for NPCs, at the end of the wave 2 of the program; there is more space for counties to lobby to affect the decision of counties' graduation. The really poor counties are more likely to be graduated because they may not have enough resources like relationships with central and provincial government leaders or money to lobby. Third, for counties in group 3 which have received the external aid for 15 years, they may have formed the dependence on governments' disbursement. The higher dependency degree on

[40] The criteria of graduation from wave 1 is that net income per capita exceeded 700 RMB in 1992; while the criteria of graduation from wave 2 is not only based on net income per capita, but also local GDP per capita, poverty headcount, and local fiscal revenue per capita. More details regarding the criteria for the designation of poor counties are available upon reqeust.

outside funds of poor counties have, the less are likely to improve the efficiency of resources. Thus, the positive impacts can only sustain three years for group 3 counties after their graduation. Last but not least, even though the received per capita funds from the program were similar for counties in each wave, the proportion of the received funds over net income per capita were 12.8%, 7.7%, and 6.6% respectively for counties in each wave. The results imply that government should reduce the number of NPCs but increase the amount of funds for each poor county and then each person in poor counties.

1.5.3 County-Level Analysis

In this section, I keep decomposing the long-term impacts of anti-poverty program to the county level. Since the results of overall average and group averages are estimated based on the individual synthetic control results, I demonstrate the details of the result of one treated county first. The three treatment groups have different control groups which range from 68 to 170 counties for each case's donor pool. The donor pool for graduated NPCs in the first treatment group are counties that participated in the first wave of the program in 1986 and stayed through the second wave; for the second group, donor units are counties that entered the second wave of the program in 1994 and stayed through the third wave; and my donor units for the third group are counties that have been designated as NPCs in 1986, and stayed in the program for 25 years, having never left.

The units in the donor pool of synthetic control are all NPCs for each sub-group. When I use the control to predict Jingtaixian's rural net income per capita from 1986 to 2000, the root mean squared percentage error (RMSPE) is 0.06, or 6 percent. A low score indicates that the synthetic control fits the pre-treatment outcome variable accurately. The post-treatment RMSPE for Jingtaixian from 2001 to 2010 is 0.24, and the ratio of post-treatment to pre-treatment RMSPE is 4.03, which is ranked as 2 out of 74, then the p-value for this

estimate is 0.027 (2/74), or 2.7%, which implies that the estimated treatment effect is not simply due to randomness.

There is a total of 73 counties in the Jingtaixian donor pool. Comparing the synthetic control to the original county in both the pre-treatment and post-treatment, I find that the average RMSPE of placebos from 1986 to 2000 is 0.12, which is higher than the RMSPE of Jingtaixian, the average RMSPE of placebos from 2001 to 2010 is 0.13, and the average ratio of RMSPE is 1.14, which indicates that the treated unit has a higher treatment effect than the placebos.

Figure 9 shows the individual results by each group. I present all cases whose pre-treatment rural net income per capita trajectories are matched with a synthetic control group. Each figure graphs the poor county that graduated solid lines and displays their synthetic counterparts with dash lines. Figure 10 presents the placebo tests that further strengthen my findings. In each panel, the bold line depicts the treated county and the thinner lines the placebos. The vertical axis represents the difference in the rural net income per capita between the treated (or the placebo) and its synthetic control. I identify a significant positive effect if the thick curve is higher than most thin lines post-treatment and an adverse effect if lower. I also calculate RMSPE and p-value for each case as in Table 5.

The results for graduated NPCs in group 1 are reported in Figure 9.A (synthetic control results) and Figure 10.A (placebo tests). Luoyuanxian, Anxixian, Pinghexian, and Linquxian are examples of a positive, robust long-run impact of anti-poverty program. For example, seven years after graduation, net income per capita is about 36% higher than that of the constructed counterfactual in Luoyuanxian, 52% in Anxixian, 71% in Pinghexian, and 68% in Linquxian. The placebo tests show that the synthetic control results are robust. The effect in the treated county is above all of the placebos. However, Gaolanxian and

Zhengningxian have lower income level relative to the synthetic control one to seven years after graduation, and none of them is robust to placebo test. Yiyuanxian provides a mixed empirical result with lower net income per capita relative to the synthetic control in the first four years after graduation and then rebounding to a higher level relative to the counterfactual. That rebound to a positive effect in five years weakens the causal link to anti-poverty program.

The synthetic control and placebo test results for group 2 are summarized in Figure 9.B and Figure 10.B. Seven counties (Xifengxian, Xiaanxian, Yichuanxian, Mianchixian, Qinyuanxian, Tengchongxian, and Zhongxian) fare better than their synthetic controls, with the range of 8% to 24% and 14% to 41%, lower than the percentages (36% to 71%) in group 1, higher net income per capita relative to that of counterfactuals seven and ten years after graduation. Only the positive impact of Qinyuanxian is robust to placebo tests. Kedongxian, Mingshuixian, and Qinggangxian are examples of long-run negative effect with 46%, 55%, and 37% lower income level than their synthetic controls seven years after graduation, and 60%, 42%, and 30% lower in ten years. The above effects are all robust to the placebo tests. Moreover, Wenxian, Xiaxian, and Xiangyunxian are examples of negative effects, but without strong support from the placebo tests. The rest of seven counties (Fenggangxian, Tunchangxian, Yixian, Wanrongxian, Moudingxian, Binchuanxian, and Shipingxian) have mixed impacts with either from positive to negative or from negative to positive impacts.

The individual results for graduated NPCs in group 3 are presented in Figure 9.C (synthetic control results) and Figure 10.C (placebo tests). Luoshanxian and Lantianxian are examples of positive impact after graduation. Net income per capita is 9% and 29% higher than synthetic controls for Luoshanxian and Lantianxian ten years after graduation, but only the positive impact of Lantianxian is robust to the placebo testing results. Net income per

capita of Jingtaixian and Yongdengxian are both 28% lower than that of constructed counterfactuals ten years after graduation. Those negative effects are both robust to placebo tests. Jingyuanxian, Qinxian, Fuhaixian, and Muleihasakezizhixian have mixed impacts after graduation and all rebound from positive to negative effects.

Summing up, most of the graduate counties fare no worse than their synthetic controls after graduation, only 10 out of 35 counties have negative impacts, of which only 5 (3 in group 2 and 2 in group 3) are robust to their placebo tests. Moreover, if I only consider the significant cases which are significant at the 10% level in Table 5, there are only four (2 in both group 2 and 3) adverse effect cases.

The non-negative impacts imply that counties are generally able to sustain growth in net income per capita after the external aid ends and are less likely to return to poverty. Once the treated unit graduates from the program, then it would have less available capital, but it may reach higher efficiency of the limited capital after several years' accumulation in the program relative to the synthetic unit. This is possible because of the increase in growth capacity resulted from the poverty reduction program within the disbursement period. More specifically, the increased local public goods like education, infrastructures, irrigation and farmland water conservancy project and health system, which the poverty reduction program ostensibly improves, attribute to the improvement in growth capacity. Chen et al. (2009) explain that the increased local public goods create a higher marginal product of private capital, and attract more private investment supported by increased savings rate from private sector under the assumption of imperfect capital market in poor counties.[41] In this case, the net income would raise with the increased local public goods through a higher marginal

[41] Figur 2 shows that the treated group has higher average resident savings balance, resident savings balance per capita and the ratio of resident savings balance/GDP. The mean value of resident savings balance per capita and the ratio of resident savings balance/GDP increased from 1993 to 2000.

product of investment even after the external disbursement ended. If the graduate counties' governments implement the poverty reduction program inefficiently, for instance, the local governments allocated the poverty funds to revenue-producing projects rather than other growth-oriented activities (Rozelle et al, 1998), the precautious saving will decrease with the cessation of external disbursements and then it will lead to failing economic performance, relative to the counties that still receive external aid.

1.5.4 Alternative Counterfactual Robustness

I use counties that never participate in the program as an alternative control group to contruct synthetic counterfactuals for graduated NPCs. Though the new control group with characteristics in pre-tretment are less similar to the graduated NPCs as the ungraduated NPCs, they can also provide comparisions for the treated group if the conctructed synthetic controls closely match the treated counties. In this case, the interpolation bias is hard to avoid. Figure 11 provide the average effects for the three groups respectively. The results of average effects for the three groups are similar to the results of original control group.42 In the long-term, graduated NPCs in group 1, on average, have positive impacts; while, graduated NPCs in group 2 and group 3 averagely experience negative effect.

1.6 Conclusion

The success of poverty reduction in China has attracted worldwide attention. Several Asian and African countries hope to replicate the success of China in reducing poverty. Before other countries replicate China's poverty reduction program, policy-makers must fully understand the long- and short-term causal effects of the program. Specifically, it is

[42] The individual synthetic control results are similar to the individual results of the original regressions in the previous section. The graphs for the individual results are available upon request.

imperative to determine whether poverty reduction program generates sustainable, long-lasting development. Although a number of papers have measured the short-term effects of China's national poverty reduction program, none analyzed the post-graduation performance of the recipient counties. Positive long-term effects are one of the substantial goals of the program and play an important role in reducing poverty in the long-run.

This article fills the gap by using synthetic control to find whether poor counties sustain income growth following the graduation or return to poverty. I estimate the long-term effects of the first two waves of poverty reduction program case by case from 1986 to 2010. I also follow Cavallo et al. (2013) to estimate the overall average effects for all treated counties and sub-groups. I find that China's anti-poverty program had a nonnegative impact on net income per capita, at least. However, I find a lot of heterogeneity in the results across counties and time. Only five graduate counties show that they were unable to keep the pace of growth as they would have if they had not left the program, and the results are robust to their placebo tests. Heterogeneity in impacts implies that poverty reduction program does not have a long-term positive effect for every poor county as they were estimated in short-term average. In groups, only counties that graduated in 1993 are, on average, able to have a higher net income per capita than they witnessed in the counterfactuals. Counties that graduated in 2000 performed better than or at least the same as the synthetic units, on average, only for three years after graduation. The results imply that only the first wave of the poverty reduction program has a demonstrable, positive, lasting impact on net income per capita for poor counties that graduated in 1993; the nonnegative impacts of the second wave of the program only last for three years after poor counties' graduation in 2000.

Constrained by a bad fit or data unavailability, I cannot analyze all treated counties. If a county is omitted due to poor fit, the above results still present unbiased estimates of

the poverty reduction program's causal effect. If a county is omitted due to data limitation, then the overall average effect would suffer from upward bias, because the treated counties without sufficient data tend to be poorer. Though the overall average effects may be biased due to missing treated counties, the synthetic control for individual case studies still captures the causal effect of poverty reduction program in China in the long-run. Future research on China's poverty reduction program should pay attention to analyze why certain counties experience negative long-run effect; test whether the dependence on external aids exists in poor counties which have benefited from the outside funds several years but remained poor; and identify methods to improve the long-run effects of poverty reduction program.

Table 1. Three sub-groups of the treated group

	Entering time	Graduating time	Sample size	Designated status during the three waves of the program (1-2-3)
Group 1	1986	1993	20	P-G-G
Group 2	1994	2000	27	NP-P-G
Group 3	1986	2000	14	P-P-G

Notes: Designated status for group1, P-G-G, means that the status of graduated NPCs in group 1 is participated during the first wave, and graduated in the second and third wave. Similarly, the status of graduated NPCs in group 2 is not participated during the first wave, participated in the second wave, and graduated in the third wave; the status of graduated NPCs in group 3 is participated during the first and second wave, and graduated in the third wave.

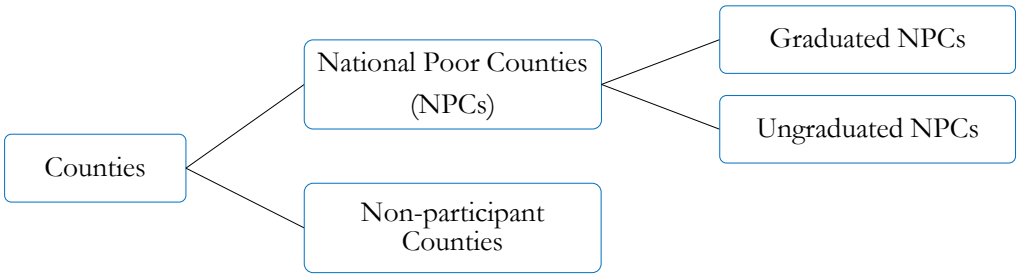
Table 2. Statistical summary

Variable	Non-participant Counties		NPCs		Ungraduated NPCs		Graduated NPCs		t-test	t-test
	Mean (1)	Std. Dev.	Mean (2)	Std. Dev.	Mean (3)	Std. Dev.	Mean (4)	Std. Dev.	(1)-(2)	(3)-(4)
GDP per capita	0.74	0.81	0.43	0.68	0.40	0.65	0.62	0.76	0.31***	-0.22***
Population	47.41	269	38.60	131.2	38.22	141.21	40.93	26.82	8.81***	-2.71*
Area per capita	701	4,736	362	1,685	364	1,750	351	1,263	339***	12.85
Fisca revenue per capita	501	1,164	330	1,099	313	1,097	451	1,102	171***	-137.82***
Primary industry added value/GDP	0.35	2.06	0.37	0.16	0.38	0.15	0.32	0.19	-0.02	0.06***
Resident savings balance per capita	4,402	3,846	2,616	2,545	2,481	2,471	3,478	2,825	1786***	-997.68***
Net income per capita	2,711	1,442	1,582	882	1,482	773	2,199	1,209	1130***	-716.78***
Hilly county	0.27	0.44	0.20	0.40	0.18	0.39	0.29	0.45	0.07***	-0.11***
Border county	0.05	0.21	0.06	0.24	0.07	0.25	0.04	0.20	-0.02***	0.02***
Pastoral county	0.12	0.33	0.15	0.36	0.15	0.35	0.18	0.39	-0.03***	-0.04***
Minority county	0.23	0.42	0.42	0.49	0.46	0.50	0.23	0.42	-0.19***	0.23***
Mountainous county	0.33	0.47	0.64	0.48	0.65	0.48	0.56	0.50	-0.32***	0.09***
CPI	0.86	0.30	0.87	0.30	0.87	0.31	0.86	0.30	-0.03	0.01
National poor county (1986-1993)	0	0	0.47	0.50	0.45	0.50	0.58	0.49		
National poor county (1994-2000)	0	0	0.85	0.36	0.86	0.35	0.78	0.41		
National poor county (2001-2010)	0	0	0.86	0.35	1	0	0.03	0.18		

Sources: Calculated from data in China County (City) Social and Economic Statistical Yearbook (1996, 2000-2011), China Statistical Yearbook for Regional Economy (2000-2011), China County Statistical Yearbook (1980-1991), China rural statistical yearbook (1987-2011), China's Ethnic Statistical Yearbook (1987-2011) as well as provincial yearbooks, cities' yearbooks, and counties' yearbooks.

Notes: Number of observations vary between 13,000 and 22,675 for Non-participant counties, between 9,173 and 16,088 for NPCs, between 8,181 and 13,773 for ungraduated NPCs, and between 1,137 and 2,315 for Graduated NPCs. Non-participant counties are counties that never have been designated as the NPCs. GDP per capita, fiscal revenue per capita, citizen savings balance per capita, and net income per capita values are based on 2000 constant prices. GDP per capita (ten thousand RMB), Population (ten thousand), Land areas per capita (square kilometer per ten thousand people), Fiscal revenue per capita (RMB), Primary industry added value/GDP (%), Citizen savings balance per capita (RMB), Net income per capita (RMB). *p<0.10, **p<0.05 ***p<0.01.

Table 3. Classification of counties and effects



	Intervention	Treated counties	Comparison	Estimation
Short-run effect	Participating in the program	NPCs	Non-participant Counties	Park et al. (2002); Meng (2013); Qin and Chong (2016); Figure 3
Long-run effect	Participating in the program and then graduating from it	Graduated NPCs	Ungraduated NPCs	Synthetic control (section 5.1-5.3)
		Graduated NPCs	Non-participant Counties	Synthetic control (section 5.4)

Table 4. 5×5 Transition matrix for relative income level ranks of NPCs from 1986 to 1992

		1992				
		Bottom	Second	Third	Fourth	Top
1986	Bottom	0.74	0.16	0.05	0	0.05
	Second	0.15	0.45	0.1	0.1	0.2
	Third	0.11	0.17	0.33	0.11	0.28
	Fourth	0	0.21	0.42	0.26	0.11
	Top	0	0	0.11	0.56	0.33

Notes: This table reports the probabilities of net income per capita ranks transition for NPCs from 1986 to 1992. The bottom, second, third, fourth, top in the quintile matrix measure the percentile of 20^{th}, 40^{th}, 60^{th}, 80^{th} and 100^{th} respectively for the income level groups.

Table 5. RMSPE for placebo test

Treated county	Pre-treatment RMSPE	Post-treatment RMSPE	Ratio	Number of donors	Placebo average pre-treatment RMSPE	Placebo average post-treatment RMSPE	Average ratio	rank	p-value	Effect direction
Group 1: 1986-1993										
Luoyuanxian	0.09	0.2	2.14	78	0.07	0.18	4.07	46	0.58	+
Anxixian	0.06	0.26	4.34	78	0.07	0.18	3.85	28	0.35	+
Pinghexian	0.07	0.28	4.09	78	0.07	0.19	3.91	28	0.35	+
Gaolanxian	0.05	0.26	5.29	68	0.08	0.2	3.46	16	0.23	-
Zhengningxian	0.02	0.01	4.06	68	0.08	0.2	3.28	20	0.29	-
Linquxian	0.08	0.3	3.89	69	0.07	0.2	4.28	27	0.39	+
Yiyuanxian	0.08	0.12	1.62	69	0.07	0.2	4.23	50	0.71	0
Group 1 average										+
Group 2: 1994-2000										
Xifengxian	0.04	0.09	2.33	153	0.07	0.17	3.5	71	0.46	+
Fenggangxian	0.13	0.15	1.22	153	0.07	0.17	3.43	115	0.75	0
Tunchangxian	0.1	0.14	1.43	153	0.07	0.16	3.42	108	0.7	0
Yixian	0.05	0.06	1.26	153	0.07	0.16	3.45	112	0.73	0
Kedongxian	0.1	1.12	11.14	154	0.07	0.17	3.48	10	0.06	-
Mingshuixian	0.09	0.81	9.31	154	0.07	0.17	3.44	11	0.07	-
Qinggangxian	0.19	0.52	2.7	154	0.07	0.17	3.54	64	0.41	-
Xinanxian	0.02	0.08	4.36	154	0.07	0.17	3.51	32	0.21	+
Yichuanxian	0.03	0.09	2.82	154	0.07	0.17	3.44	63	0.41	+
Mianchixian	0.04	0.07	1.64	154	0.07	0.17	3.48	99	0.64	+
Qinyuanxian	0.06	0.23	3.6	154	0.07	0.17	3.46	102	0.66	+
Wanrongxian	0.06	0.14	2.29	154	0.07	0.16	3.33	71	0.46	0
Wenxian	0.06	0.2	3.54	154	0.07	0.17	3.44	45	0.29	-
Xiaxian	0.06	0.22	3.69	154	0.07	0.17	3.35	39	0.25	-
Tengchongxian	0.01	0.14	9.93	154	0.07	0.17	3.49	11	0.07	+
Moudingxian	0.08	0.03	0.44	154	0.07	0.17	3.43	150	0.97	0
Binchuanxian	0.03	0.18	5.63	154	0.07	0.16	3.57	22	0.14	0
Xiangyunxian	0.03	0.08	3.04	154	0.07	0.17	3.46	59	0.38	-
Shipingxian	0.14	0.11	0.77	154	0.07	0.17	3.35	128	0.83	0
Zhongxian	0.02	0.1	6.7	170	0.05	0.16	6.01	34	0.2	+
Group 2 average										-
Group 3: 1986-2000										
Jingtaixian	0.06	0.24	4.03	73	0.12	0.13	1.14	2	0.03	-
Jingyunxian	0.03	0.17	5.38	73	0.12	0.13	1.13	1	0.01	0
Yongdengxian	0.1	0.29	3.1	73	0.12	0.13	1.1	5	0.07	-
Luoshanxian	0.03	0.05	2.02	129	0.08	0.13	2.41	51	0.39	+
Qinxian	0.13	0.38	2.83	72	0.13	0.14	1.12	5	0.07	0
Lantianxian	0.15	0.17	1.14	108	0.09	0.12	1.67	59	0.54	+
Fuhaixian	0.04	0.09	1.99	130	0.05	0.13	4.75	90	0.69	0

Muleihasakezizhixian	0.07	0.12	1.57	162	0.05	0.12	5.21	119	0.73	0
Group 3 average										-

Notes: "+" indicates positive impact, which means that the treated values are larger than the synthetic values after the poor counties graduate from the program for most years; "-" indicates negative impact, which means that the treated values are smaller than the synthetic values after the poor counties graduate the program for most years; "0" indicates mixed (no obvious positive or negative) impacts , which means that the treated values are sometimes larger than and sometimes smaller than the synthetic values after the graduated NPCs leave the program. a. The pre-treatment RMSPE is very high for the synthetic of Kedongxian in Heilongjiang province, so I exclude out this case from group 2.

Figure 1. Count of counties by classification (1986-2010)

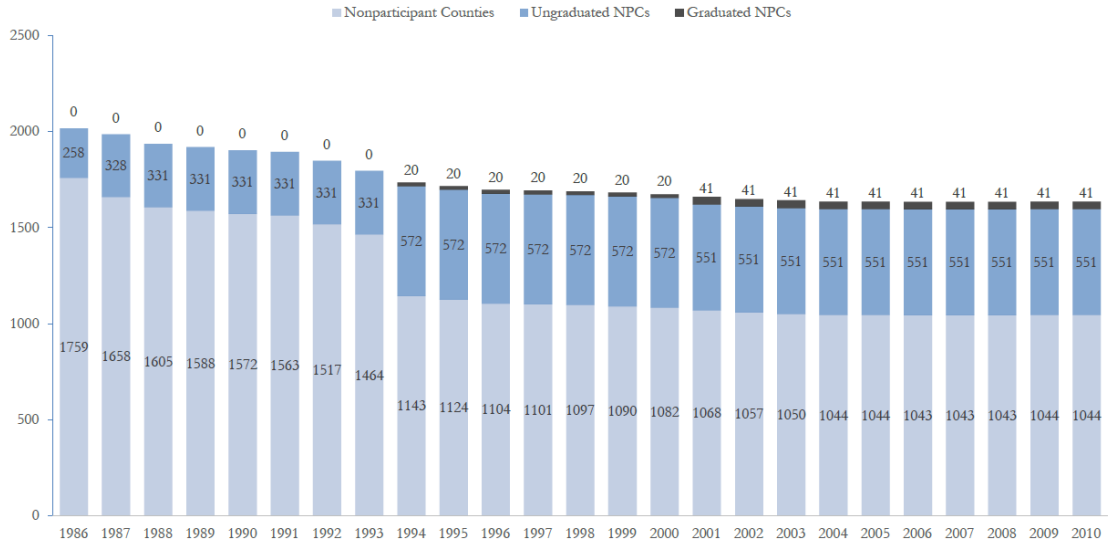


Figure 2. County map of China by NPCs status by wave

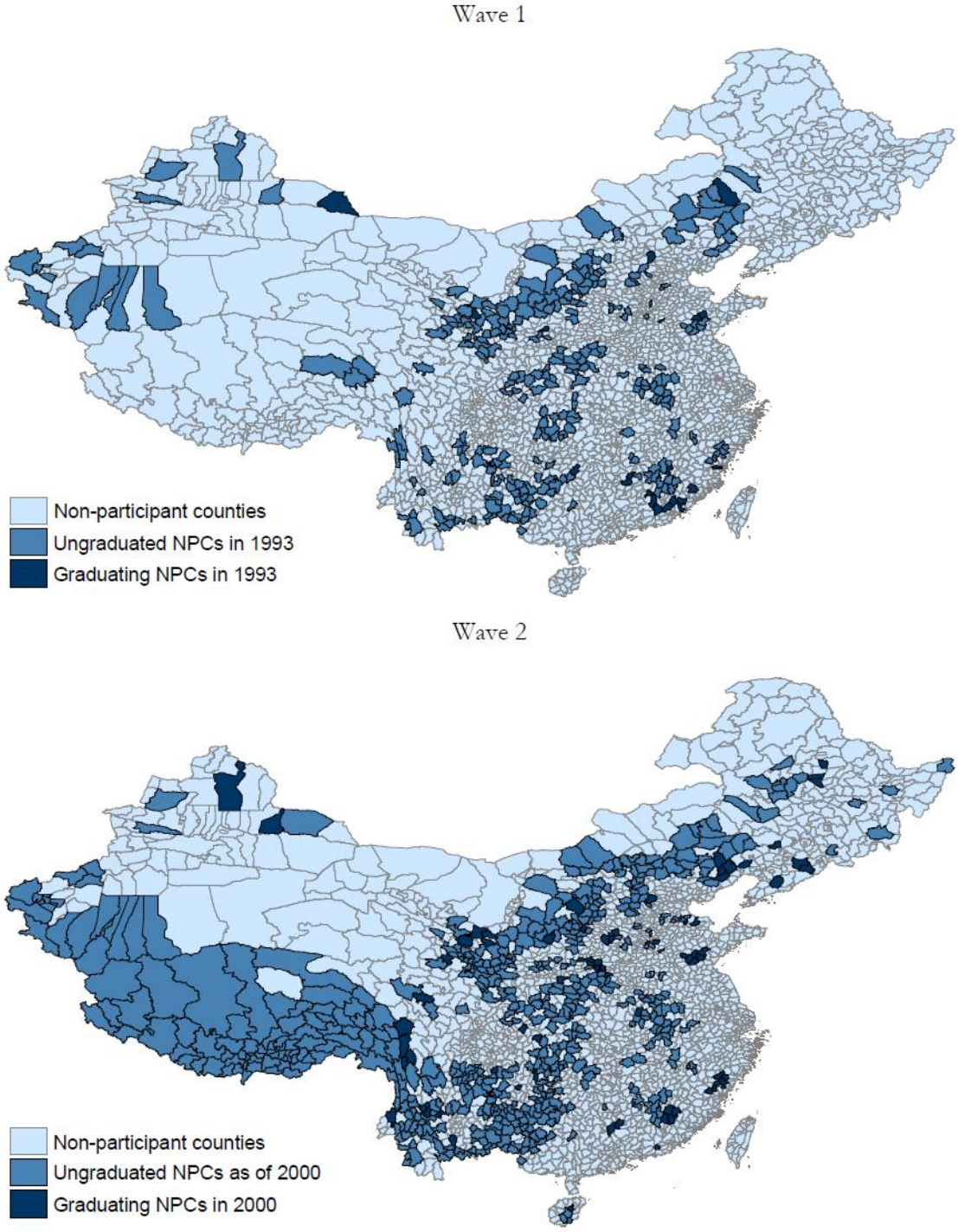
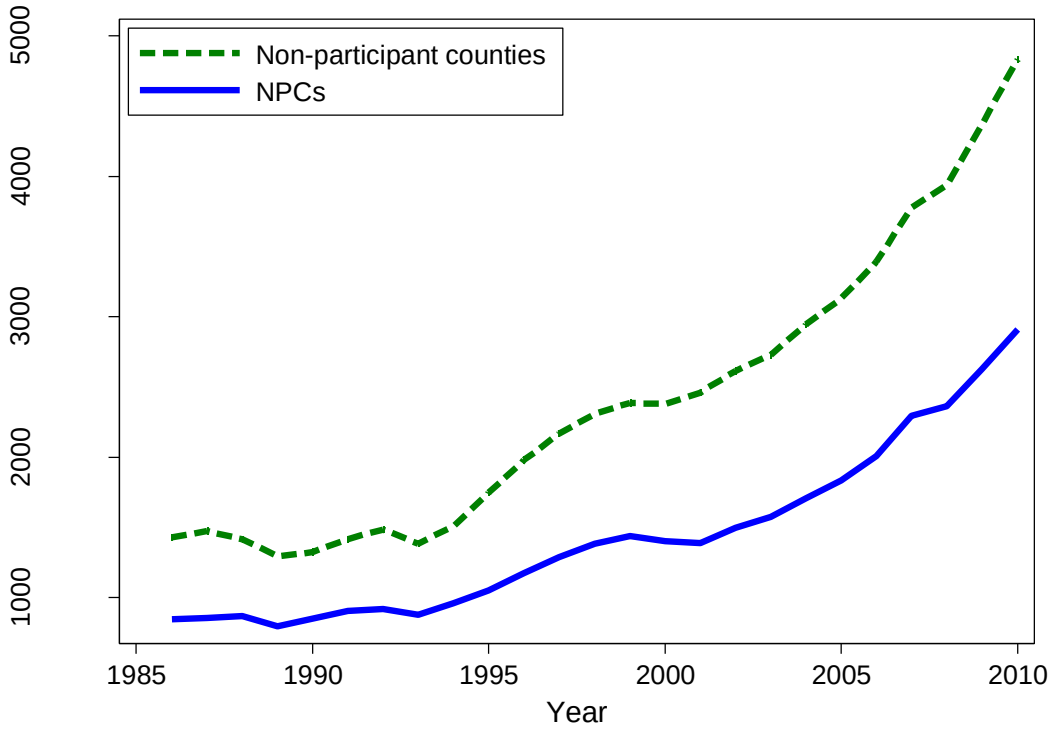
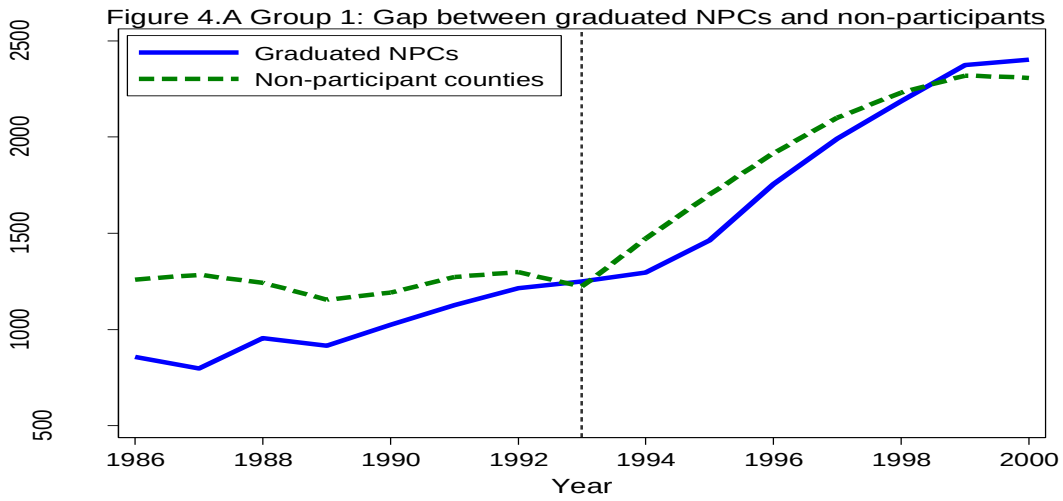


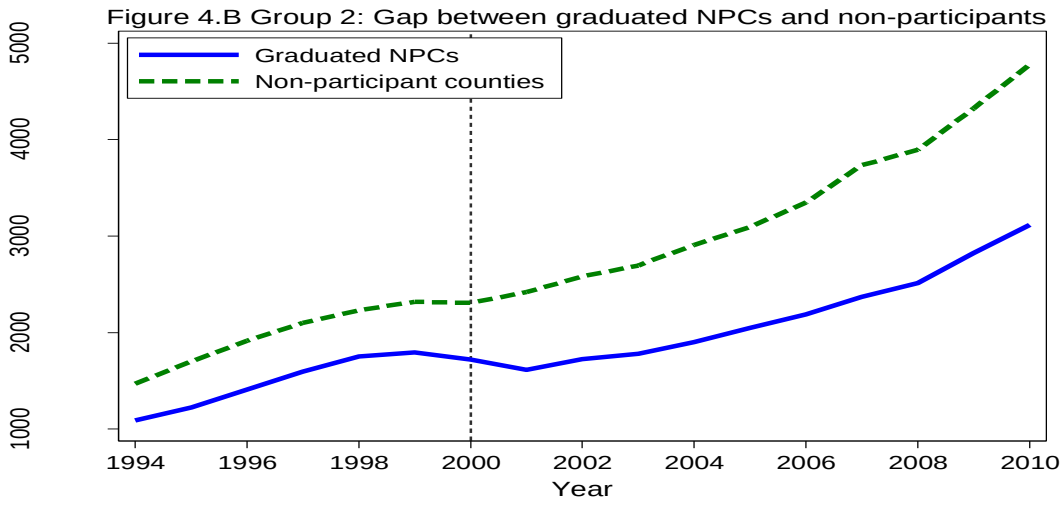
Figure 3. Rural net income per capital from 1986 to 2010



Notes: Non-participant counties are counties that never have been designated as NPCs. Graduated NPCs are those NPCs that graduated from the poverty reduction program.

Figure 4. Gap between graduated NPCs and non-participants





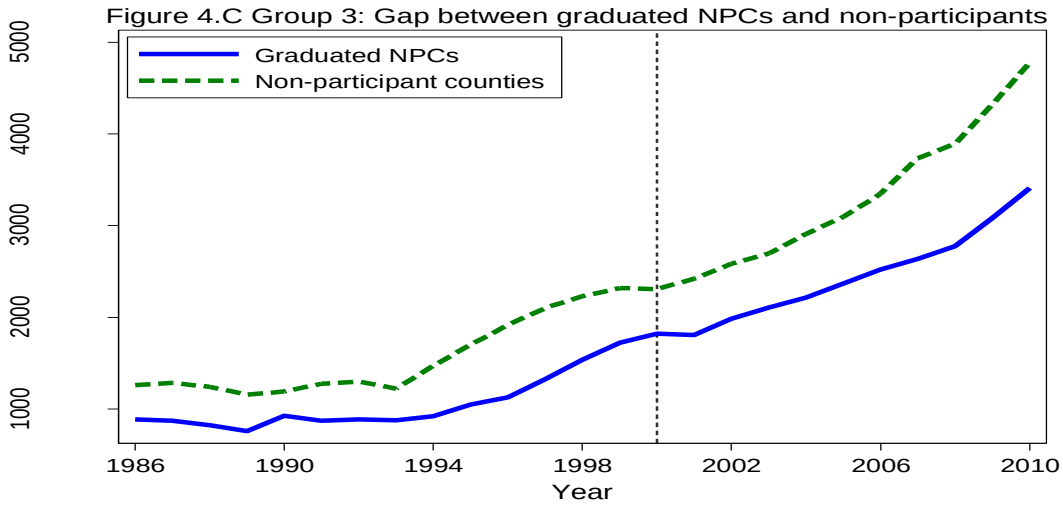


Figure 5. Estimated average long-term effects in post-graduation for full sample

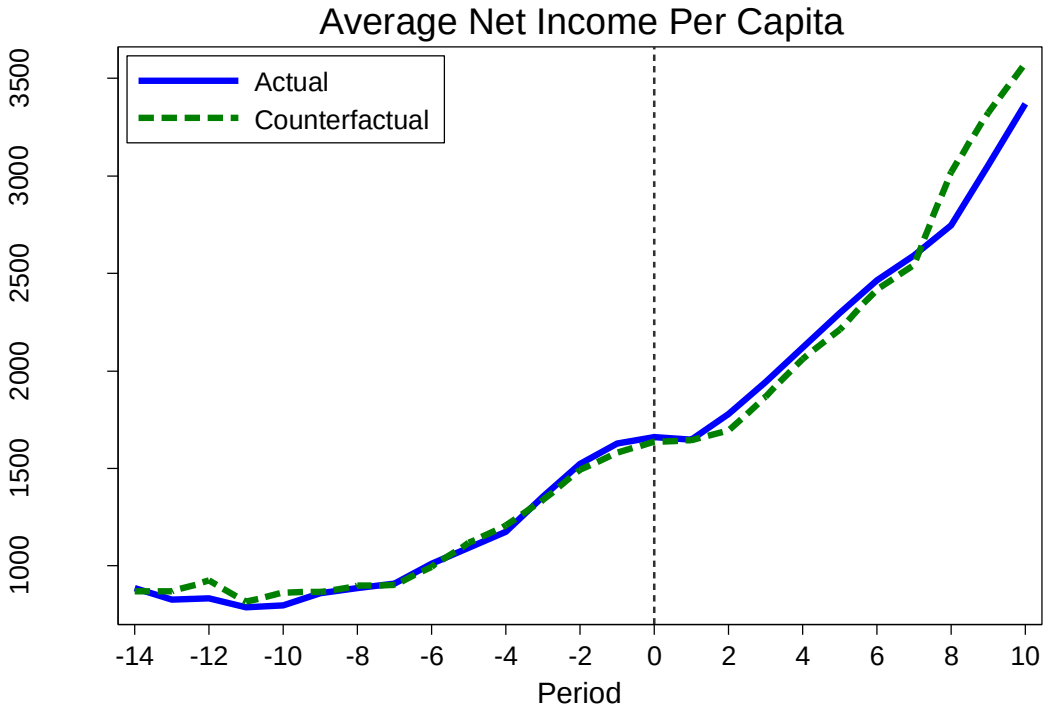
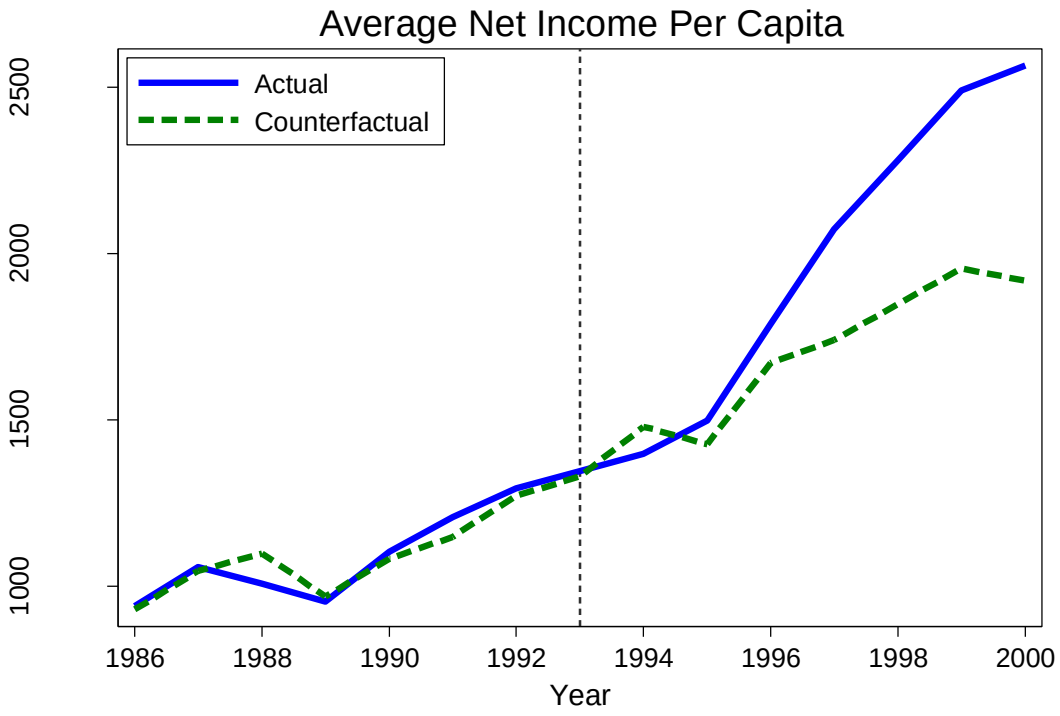


Figure 6. Average net income per capita and significance levels for graduated NPCs in group 1 (1986-2000)



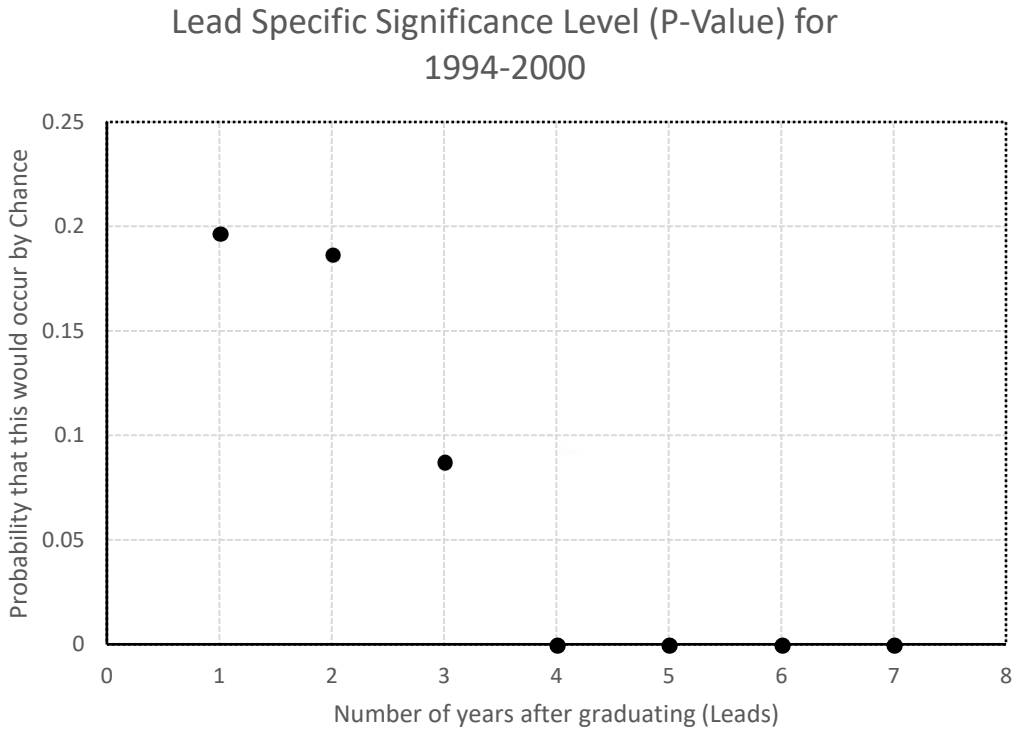
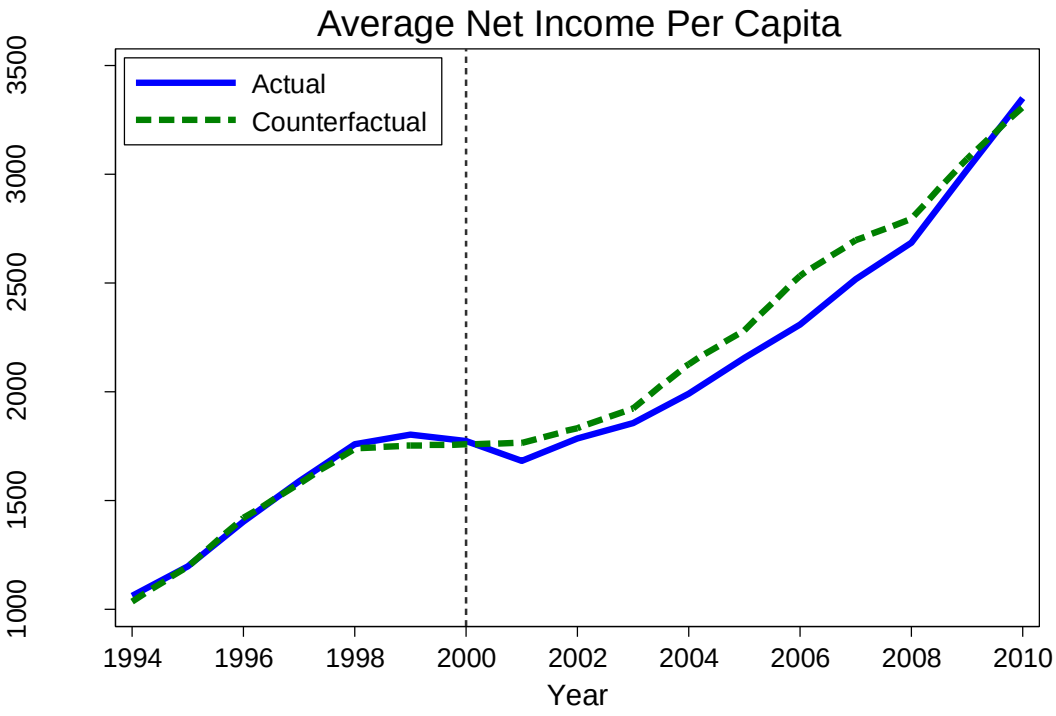


Figure 7. Average net income per capita and significance levels for graduated NPCs in group 2 (1994-2010)



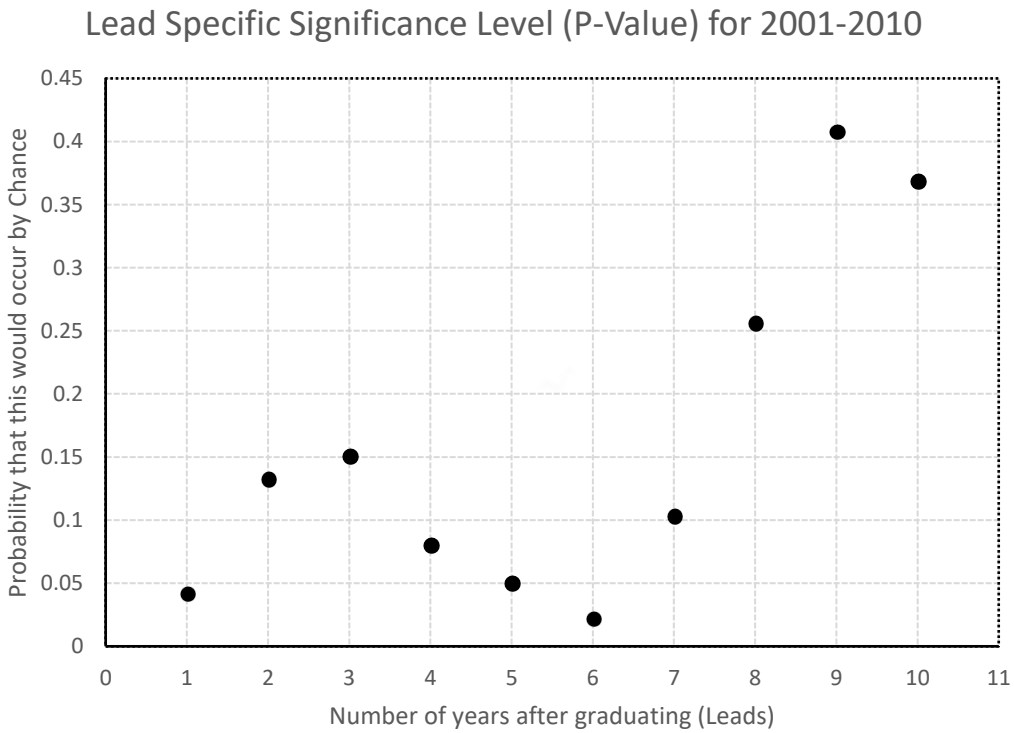
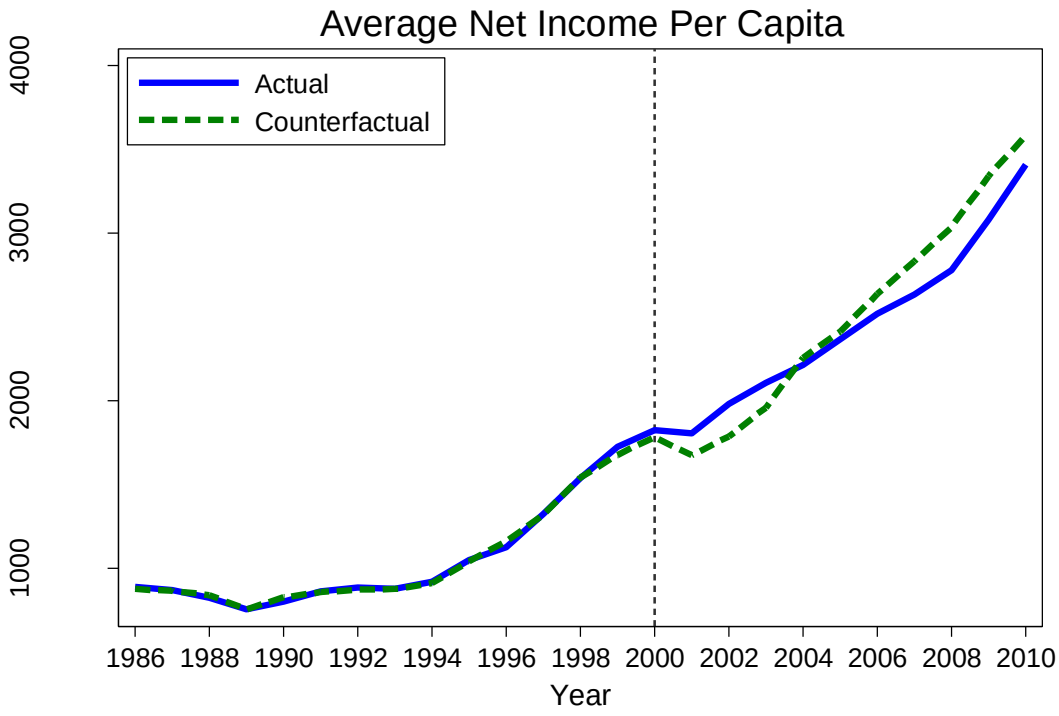


Figure 8. Average net income per capita and significance levels for graduated NPCs in group 3 (1986-2010)



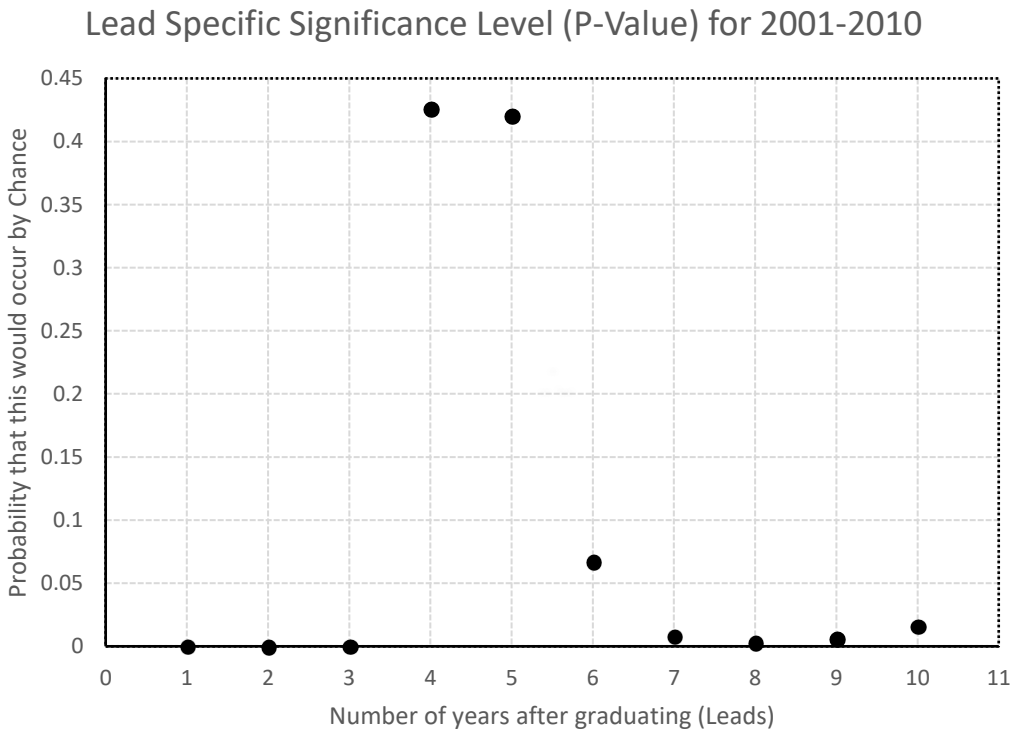
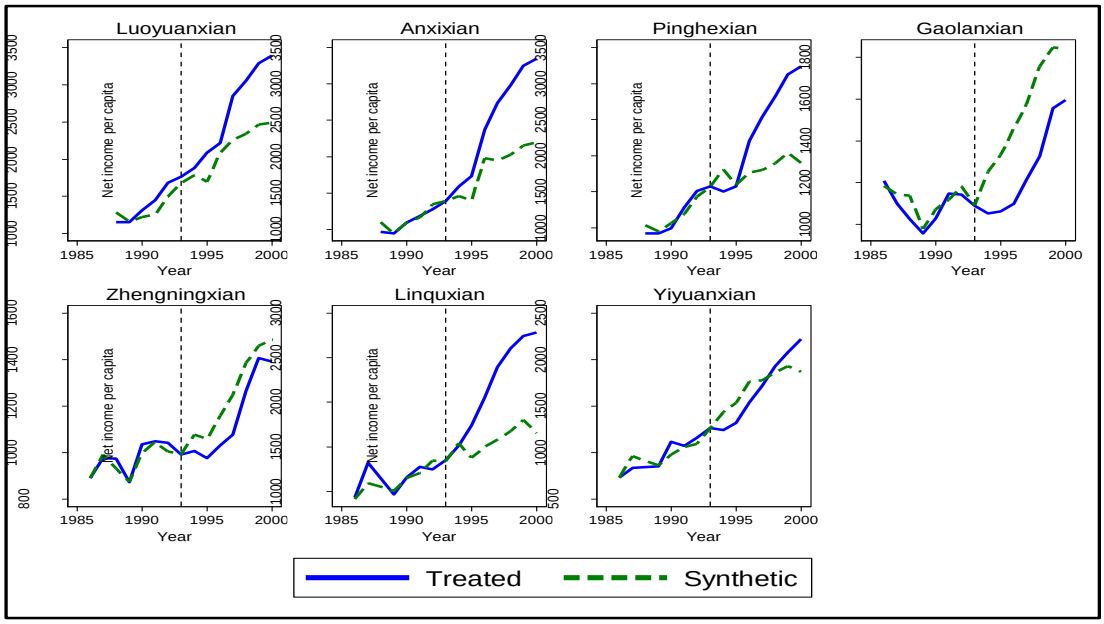
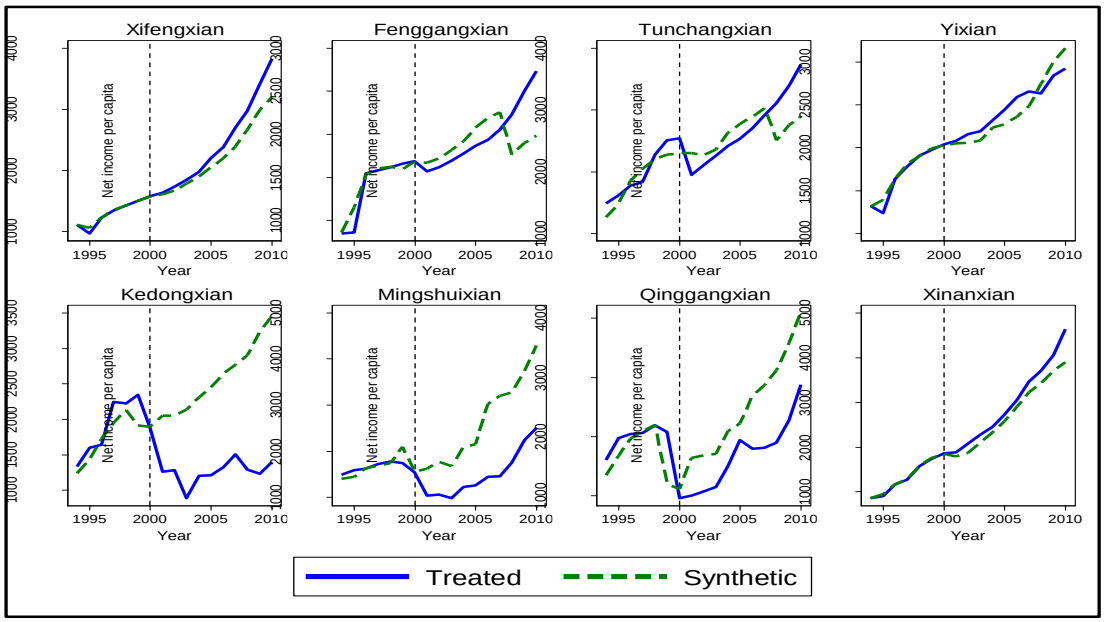


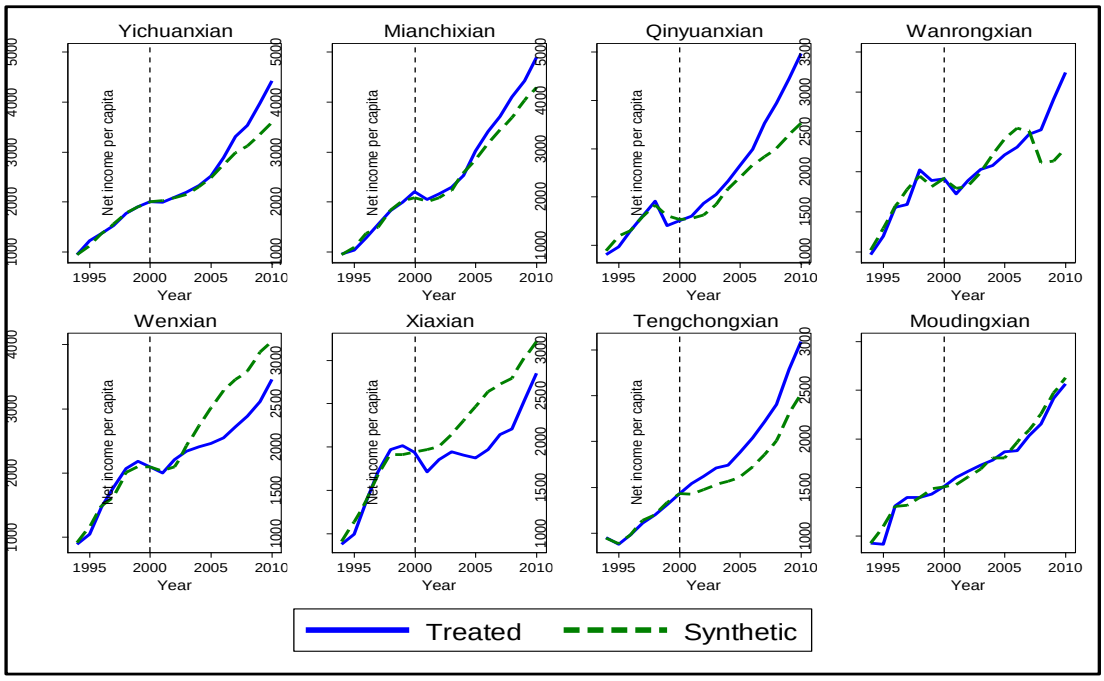
Figure 9. Synthetic control results for counties participated in the poverty reduction program

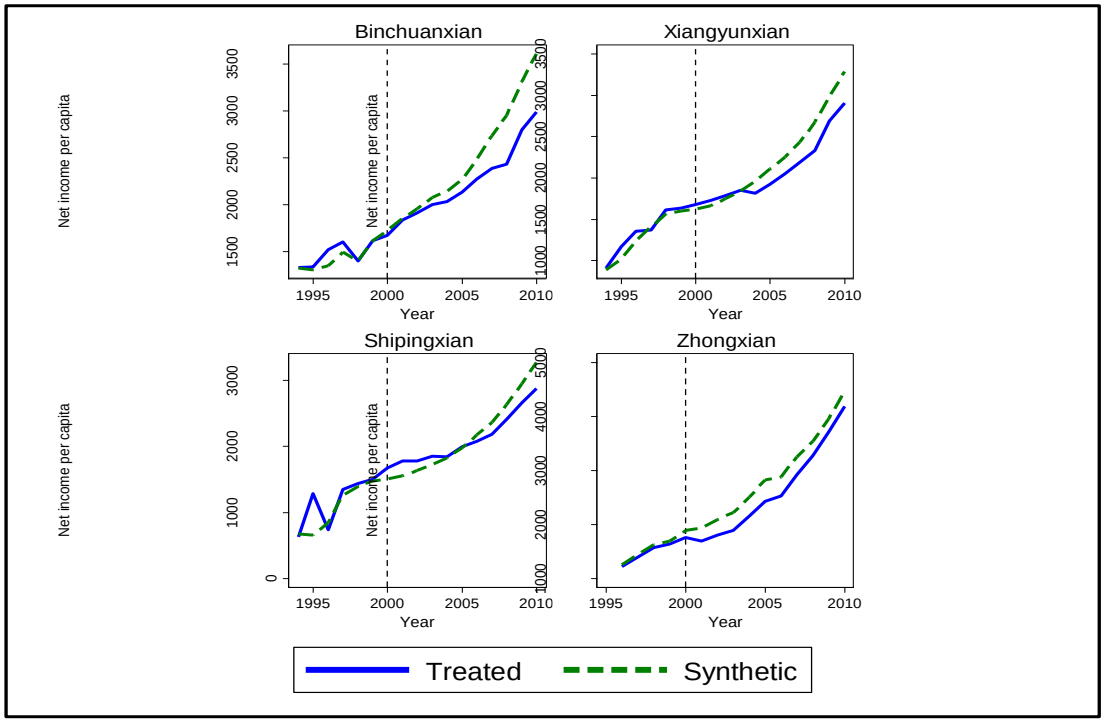
Panel A Synthetic control results for counties participated in the poverty reduction program in 1986 and graduated in 1993 (group 1)



Panel B Synthetic control results for counties participated in the poverty reduction program in1994 and graduated in 2000 (group 2)




Yichuanxian
Mianchixian
Qinyuanxian
Wanrongxian
Wenxian
Xiaxian
Tengchongxian
Moudingxian
Net income per capita
Year
Treated
Synthetic


Binchuanxian
Xiangyunxian
Shipingxian
Zhongxian
Net income per capita
Year
Treated
Synthetic

C Synthetic control results for counties participated in the poverty reduction program in1986 and graduated in 2000 (group 3)

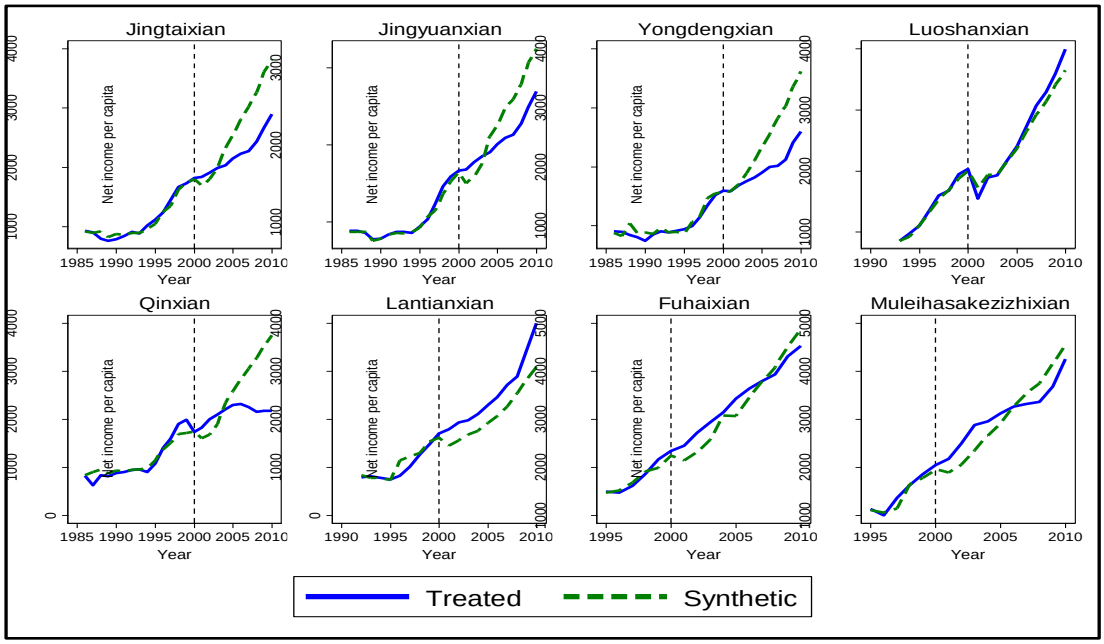
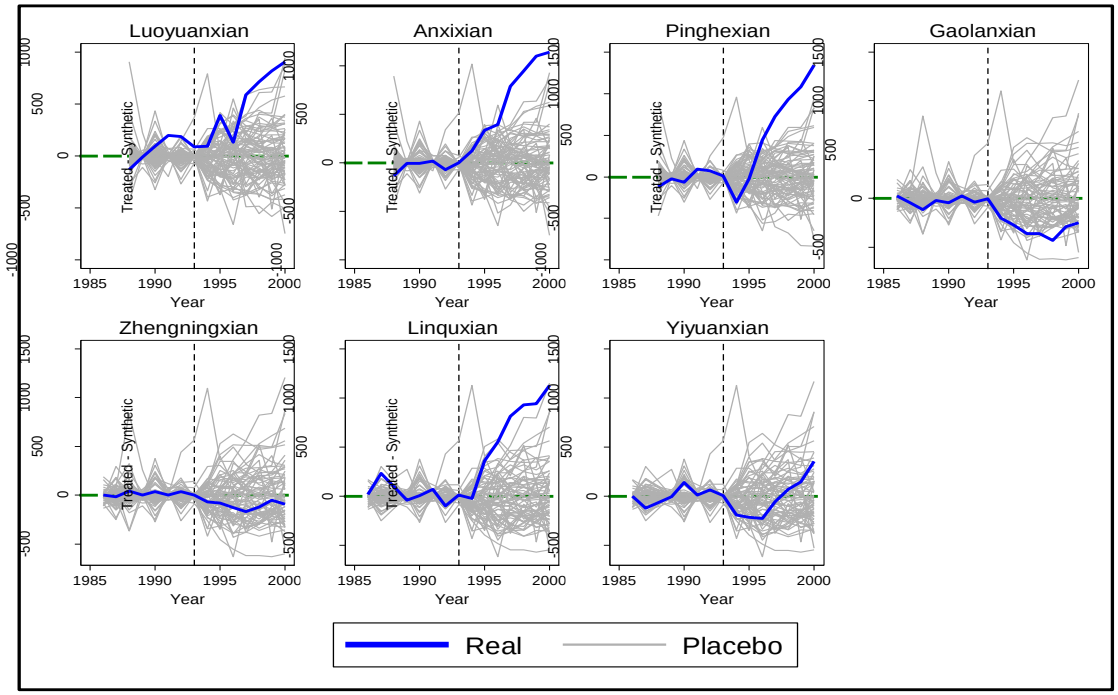
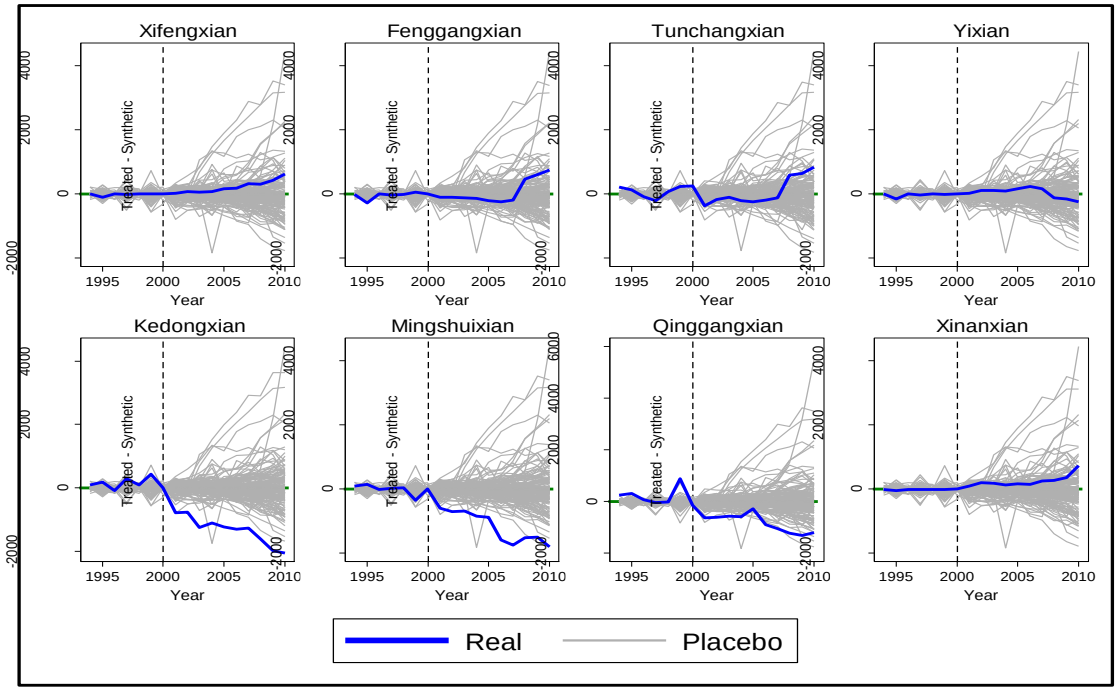


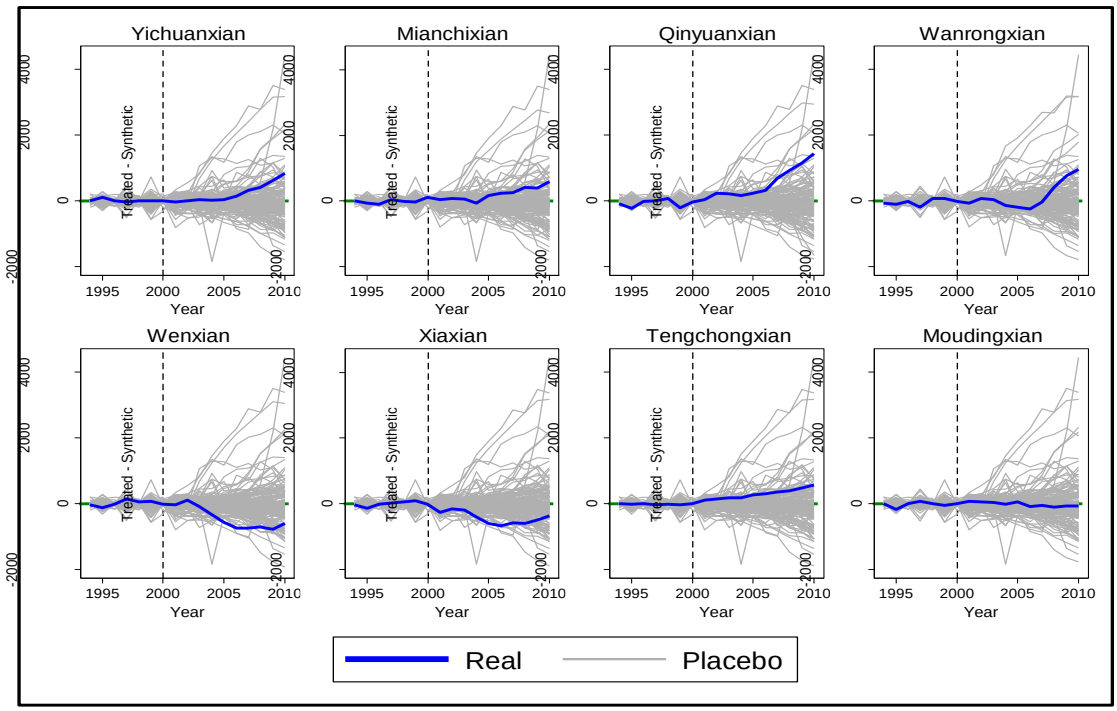
Figure 10. Placebo test results for counties participated in the poverty reduction program

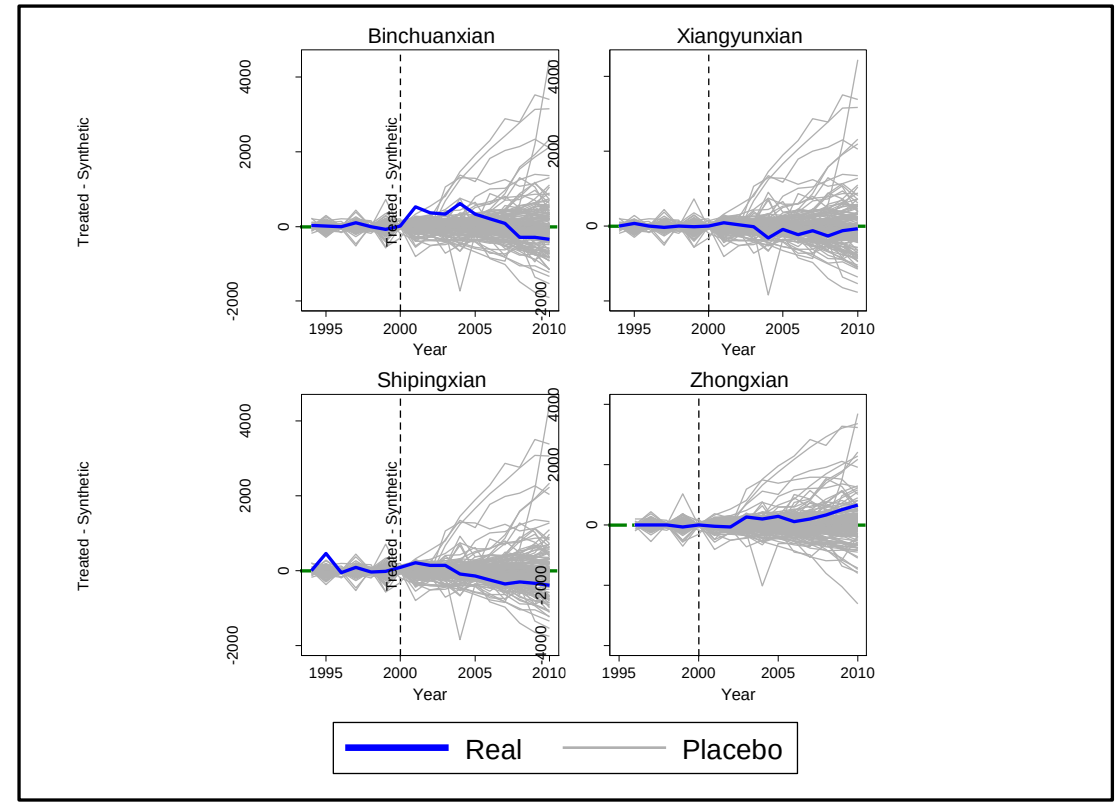
Panel A Placebo test results for counties participated in the poverty reduction program in1986 and graduated in 1993 (group 1)



Panel B Placebo test results for counties participated in the poverty reduction program in1994 and graduated in 2000 (group 2)




Yichuanxian
Mianchixian
Qinyuanxian
Wanrongxian
Wenxian
Xiaxian
Tengchongxian
Moudingxian
Treated - Synthetic
Year
Real
Placebo


Binchuanxian
Xiangyunxian
Shipingxian
Zhongxian
Treated - Synthetic
Year
Real
Placebo

Panel C Placebo test results for counties participated in the poverty reduction program in1986 and graduated in 2000 (group 3)

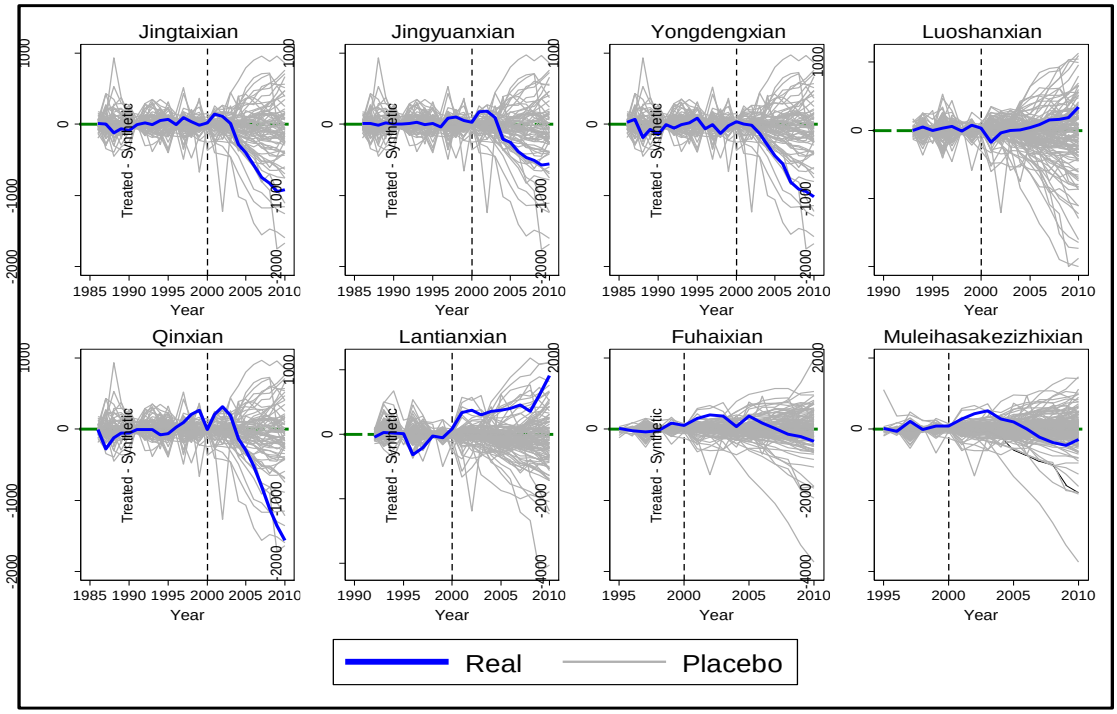
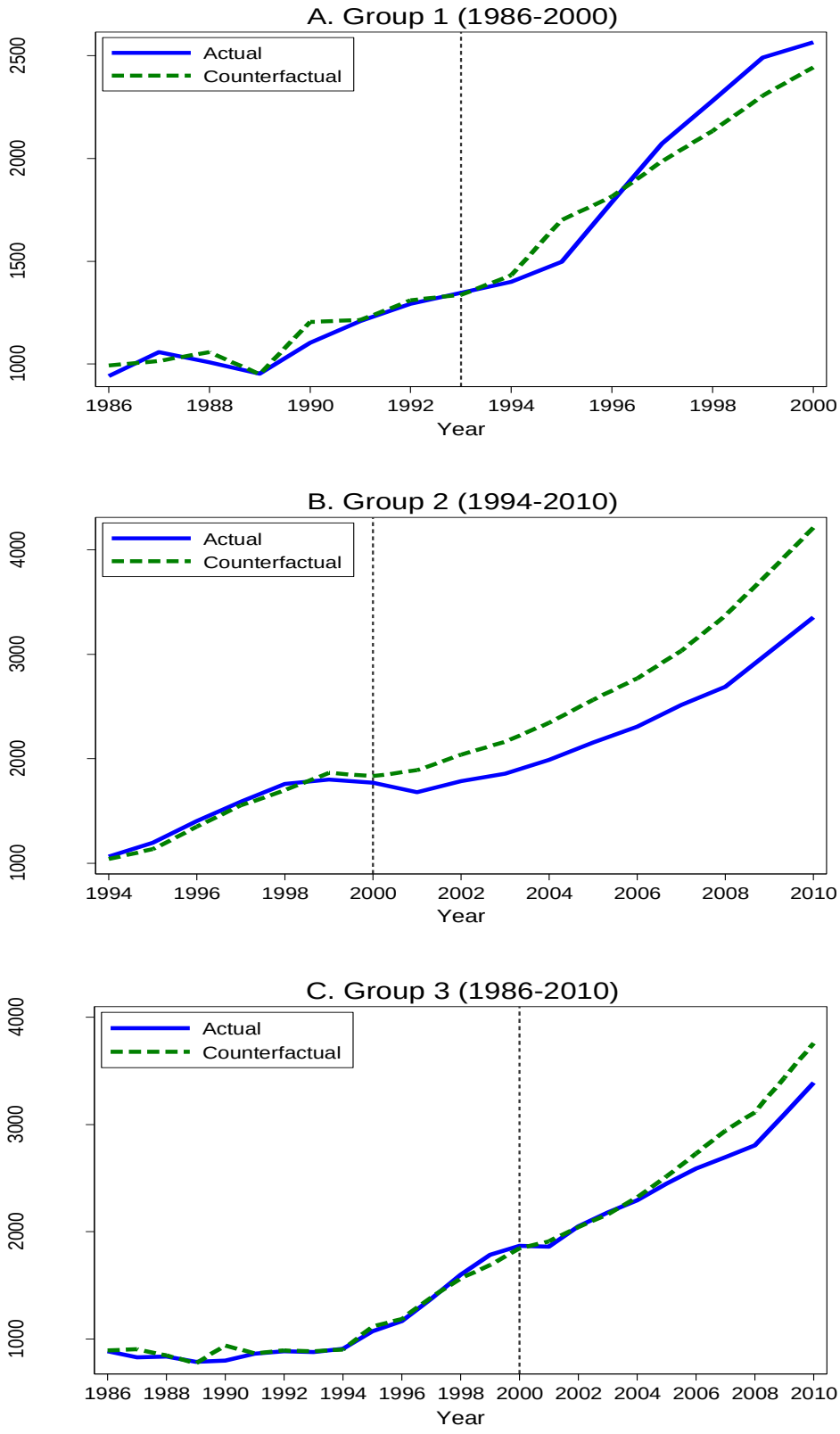


Figure 11. An alternative counterfactual



References

Abadie, Alberto, and Javier Gardeazabal. "The economic costs of conflict: A case study of the Basque Country." *American Economic Review* (2003): 113-132.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*105.490 (2010).

_____, "Comparative politics and the synthetic control method." *American Journal of Political Science* 59.2 (2015): 495-510.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. "Catastrophic natural disasters and economic growth." *Review of Economics and Statistics* 95.5 (2013): 1549-1561.

Chen, Shaohua, and Martin Ravallion. "The developing world is poorer than we thought, but no less successful in the fight against poverty." *Quarterly Journal of Economics* 125.4 (2010).

Chen, Shaohua, Ren Mu, and Martin Ravallion. "Are there lasting impacts of aid to poor areas?" *Journal of Public Economics* 93.3 (2009): 512-528.

De Janvry, Alain, and Elisabeth Sadoulet. "Agricultural growth and poverty reduction: Additional evidence." *The World Bank Research Observer* 25.1 (2010): 1-20.

Dube, Arindrajit, and Ben Zipperer. *Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies*. No. 8944. Institute for the Study of Labor (IZA), 2015.

Duclos, Jean-Yves, Abdelkrim Araar, and John Giles. "Chronic and transient poverty: Measurement and estimation, with evidence from China." *Journal of Development Economics* 91.2 (2010): 266-277.

Ferreira, Francisco HG, Phillippe G. Leite, and Martin Ravallion. "Poverty reduction without economic growth?: Explaining Brazil's poverty dynamics, 1985–2004." *Journal of Development Economics* 93.1 (2010): 20-36.

Galiani, Sebastian and Quistorff, Brian. "The synth_runner package: Utilities to automate synthetic control estimation using synthe." *Unpublished. URl: http://www. oekonometrie. uni-saarland. de/papers/SCM_ Predictors. pdf* (2015).

Grier, Kevin, and Norman Maynard. "The economic consequences of Hugo Chavez: A synthetic control analysis." *Journal of Economic Behavior & Organization* 125 (2016): 1-21.

Jalan, Jyotsna, and Martin Ravallion. "Transient poverty in postreform rural China." *Journal of Comparative Economics* 26.2 (1998): 338-357.

_____,"Are there dynamic gains from a poor-area development program?" *Journal of Public Economics* 67.1 (1998): 65-85.

Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. "Synthetic control methods: Never use all pre-intervention outcomes as economic predictors." *Unpublished. URl: http://econweb.umd.edu/~galiani/files/synth_runner.pdf*.

Loayza, Norman V., and Claudio Raddatz. "The composition of growth matters for poverty alleviation." *Journal of Development Economics* 93.1 (2010): 137-151.

Meng, Lingsheng. "Evaluating China's poverty alleviation program: A regression discontinuity approach." *Journal of Public Economics* 101 (2013): 1-11.

Montalvo, Jose G., and Martin Ravallion. "The pattern of growth and poverty reduction in China." *Journal of Comparative Economics* 38.1 (2010): 2-16.

Park, Albert, and Sangui Wang. "Community-based development and poverty alleviation: An evaluation of China's poor village investment program." *Journal of Public Economics* 94.9 (2010): 790-799.

Park, Albert, Sangui Wang, and Guobao Wu. "Regional poverty targeting in China." *Journal of Public Economics* 86.1 (2002): 123-153.

Qin, Cong, and Terence Tai Leung Chong. "Can Poverty be Alleviated in China?" *Review of Income and Wealth* (2016).

Ravallion, Martin, and Shaohua Chen. "Hidden impact? Household saving in response to a poor-area development project." *Journal of public economics* 89.11 (2005): 2183-2204.

Ravallion, Martin, and Shaohua Chen. "China's (uneven) progress against poverty." *Journal of Development Economics* 82.1 (2007): 1-42.

Ravallion, Martin. "Are there lessons for Africa from China's success against poverty?" *World Development* 37.2 (2009): 303-313.

Rozelle, S., Park, A., Benziger, V., & Changqing, R."Targeted poverty investments and economic growth in China." *World Development* 26.12 (1998): 2137-2151.

Sala-i-Martin, Xavier. "The world distribution of income: falling poverty and… convergence, period." *The Quarterly Journal of Economics* 121.2 (2006): 351-397.

Sangui, W., Zhou, L., Yanshun, R., & Liu, M. F."The 8-7 national poverty reduction program in China–The national strategy and its impact." *Ponencia en la Shanghai Poverty Conference, Shanghai*. Vol. 25. 2004.

Suryahadi, Asep, Daniel Suryadarma, and Sudarno Sumarto. "The effects of location and sectoral components of economic growth on poverty: Evidence from Indonesia." *Journal of Development Economics* 89.1 (2009): 109-117.

Chapter 2. Education as Opportunity? The Causal Effect of Education on Labor Market Outcomes in Jordan

2.1 Introduction

Rapid gains in education without corresponding improvements in job prospects have been cited as one of the leading causes of the Arab Spring (Campante and Chor, 2012). Several nations across the Middle East and North Africa region have experienced remarkable expansions in educational attainment over the past three decades without correspondingly large economic gains in the labor market. The Hashemite Kingdom of Jordan is no exception. Despite having the 9th largest increase in years of education around the globe from 1980 to 2010, today the nation exhibits double-digit unemployment rates (15% for men and 33% for women), a female labor force participation rate around 15%, and a wage premium for higher education lower than that for secondary school completion. Like many of its neighbors, Jordan underwent protests during the Arab Spring in 2011 and continued weakness in the labor market remains a catalyst for new demonstrations taking place in 2018.

In this paper we analyze the labor market effects of recent educational reforms in Jordan which extended the length of compulsory schooling. Although compliance with the compulsory schooling reform was imperfect, our fuzzy regression discontinuity designs estimates suggest that the 1988 educational reform increased schooling attainment among the population obtaining an educational degree by nearly a quarter of a year in the first year, and implementation improved gradually in the years after.[43] Our estimates suggest that the additional education is responsible for a small increase in the probability of entering the

[43] The average increase in educational attainment differs depending on whether the sample includes graduates with secondary and higher degrees. While compulsory preparatory schooling was lengthened by a year, this only impacted attainment for those stopping with this as a terminal degree, as secondary school (which was non-compulsory) was concurrently shortened from 3 to 2 years.

workforce and being employed, an increase in the use of self-employment, and gains in hours worked and wages.

We disaggregate these gains by gender, showing that the impacts on Jordanians at the time of their entrance into the labor market of additional schooling are significantly larger for women, particularly on the extensive margin. Males instead realize gains in hours worked, which is consistent with the fact that nearly all men already participate in the labor market in by their mid-20s. Exploiting novel features of our data to analyze affected cohorts at later points in their life, we then show that the labor market gains realized by males appear to engage more strongly over this window and to benefit both at the intensive and extensive margin over the course of their careers.

Analysis of schooling reform in Jordan presents particularly interesting case study for a number of reasons. First, while many of the causal studies of education in the empirical economics literature have relied on extensions to compulsory schooling or changes to school leaving age laws, this research has demonstrated heterogeneous effects. Some papers have found positive, significant, and sizeable estimates of the returns to schooling (Oreopoulos, 2006, 2007), while other studies have uncovered contexts in which the return to schooling appears to have been either quite low or even zero (Albouy and Lequien, 2008; Pischke, and Wachter, 2008). These variations highlight the sizeable degree to which educational investments exhibit heterogeneous returns across settings and time. Furthermore, research has historically been focused primarily on high income settings, for example, studying reforms in the United States or Europe, which are largely secular educational systems. Less is known about how well the estimates produced in these studies generalize to regions, particularly those in the developing world, with dramatically different curricula and

educational agendas with non-standard objective functions, as well as under different sets of economic conditions and societal norms of behavior.

Our work also relates to a number of papers which have examined the causal impact of education on non-labor market outcomes. For example, studies have demonstrated an impact of increased education on the probability of voting (Milligan et al., 2004), health and mortality (Silles, 2009; Kippersluis et al., 2010; Powdthavee, 2010), teenage childbearing (Black et al., 2008; Silles, 2011), and intergenerational human capital accumulation (Oreopoulos, 2006; Assaad and Saleh, 2016).

The remainder of this paper is organized as follows. Section 2 provides the background on the economy as well as on the history of education reform in Jordan. Section 3 describes the data used in the analysis. Section 4 details our empirical strategy. Section 5 presents results for the impact of education on labor market outcomes both initially and over the life cycle. Section 6 presents a series of robustness checks regarding the empirical strategy and data assumptions. Section 7 concludes.

2.2 Background

2.2.1 The Government and Economy of Jordan

The Hashemite Kingdom of *Jordan* gained independence as a hereditary constitutional monarchy from the British in 1946. Jordan's economy in the post-independence period has been affected by regional instability and conflict, including a series of international conflicts with Israel (1948/1949 First Arab Israeli War; 1967 Six Day War; 1967-1970 War of Attrition; 1973 October War) and domestically, through confrontations with the PLO, most notably the events surrounding Black September which persisted from 1970 to 1971 (Robbins, 2004). By comparison, our study period which spans education and

labor market participation in the 1980s, 1990s, and 2000s has been one of relative stability for the county. As of the mid 2010s, the nation again finds itself front and center in a refugee crisis, with inflows from socio-economic and political turbulence in neighboring countries.

These periods of conflict have complicated provision of education in Jordan, as large refugee populations have sporadically entered Jordan. Notably, during the 1948 Palestinian Exodus – some 250,000 to 300,000 Palestinians arrived in Jordan, with smaller migration flows continuing over the period 1949-1956 and again another quarter million individuals fled the West Bank and Gaza for Jordan during the 1967 Exodus. As Jordan did not have a sizeable population of Sephardi or Mizrahi, it also did not experience a similar outmigration during the corresponding Jewish exodus from Arab countries from 1948 into the 1970s, such as that from Yemen, Libya, or Iraq. Periodic inflows of population both strained resources and acted as catalyst for the expansion of educational infrastructure in the state. Today Jordan is home to the second largest refugee population in the world on a per capita basis. These inflows, coupled with high fertility rates, have led to a rapid population expansion in the country, with some of the largest gains in the urban centers such as Amman.

Economically, Jordan is still considered a developing nation. The economy expanded rapidly from the 1970s through the mid-1980s, with GDP growth often exceeding 10% a year. Jordan has occasionally been labelled a rentier state because of its large reliance on external support. This pattern peaked in the 1970s when dramatically increased aid flows arrived from a boom in Middle East oil revenues. These began to decline in late 1980s saw a decline in culminating in a federal debt crisis. Slower growth returned in the 1990s, accelerating into the 2000s, and today, Jordan has a per capita income of roughly $4,000 PPP.

In spite of the economic progress over the period, unemployment in the country has remained relatively high.

2.2.2 Educational Reforms in Jordan

A number of elements make Jordan a novel setting in which to examine educational reform. Having only existed as a formal polity for around 70 years, the country is still in the process of defining and shaping its educational policy as well as expanding access to and resources for the educational system. While seen as a strong performer in comparison to peers in the MENA Region, Jordan is still rather unique for the sheer volume of refugee flows with which the country has had to cope. Driven by conflicts in surrounding countries, Jordan is now second only to Lebanon for refugees on a per capita basis, with particularly large Palestinian and Syrian populations.

The Educational reform law of 1952 established education as a right of every Jordanian citizen and introduced compulsory schooling for a period of 7 years (this was subsequently reduced to 6 years in the 1954/1955 school year). Education Law no. 16 of 1964 extended schooling to 9 years (known at the time as preparatory school) as compulsory and ushered in a broad expansion to education - including more schools, more teachers, increased teacher training, and a unified philosophy of education (Abbas, 2012). This reform would ultimately have to cope with a large increase in demand for schooling, given large numbers of Palestinians arriving from the West Bank into Jordan in the aftermath of the 6 day war in 1967.

In this paper, we focus on the impact of provisional Education Act No. 27 in 1988 whose primary effect was to increase free compulsory schooling from 9 to 10 years (renamed from Preparatory to Basic education). All children 6 years and older are required to attend

school and must continue through ten years of basic education (Hodges, 2005). While the previous reforms are larger increases in years of education, two things complicate the analysis of their impact on labor market outcomes. First, they were implemented gradually as a result of limited state resources for education, and second, they occurred far enough in the past that many respondents in the Census and EUS surveys affected by the reforms may have exited the labor market at the time survey.

Compulsory schooling in Jordan is equivalent in duration to completing primary school in the U.S. as well as two years of high school. After completing Basic education, students are eligible to enter secondary school. Prior to the reform, secondary school was heterogeneous in nature and lasted 3 years. After the reform, a "new secondary" program was implemented which standardized the curriculum, created specific tracks of study, and decreased the length of secondary school to 2 years. This means that the combined effect of Education Act No. 27 was to change the nature of secondary school, but not to alter the total number of years required to reach a secondary degree before and after the reform. The next significant revision to the educational system would not occur until 1994, when Educational Act No. 3 reformed the philosophy and objectives of education in the county.

2.3 Data and Strategy

We combine data from multiple sources in our analysis. Individual level educational attainment and labor force outcomes are derived both from the 2004 Jordanian Census (IPUMS, 2018) as well as from the Jordanian Employment and Unemployment Surveys (EUS) for 2006, 2007, 2008, 2009, 2011, 2013, 2014, and 2016. The 2004 Census is a nationally representative, stratified, 10%, sample of the Jordanian population, representing about half a million individuals. Estimates from the Department of Statistics report a

response rate of 95.9%. The EUS are representative at both the national and governate level, and are conducted quarterly. We employed the full set of harmonized EUS compiled and disseminated by the Economic Research Forum (OAMDI, 2017).

Table 6 presents summary statistics for our Census sample.[44] Panel A contains individual level demographics and work outcomes. At the time of enumeration for the 2004 census, the mean individual in the sample is 30 years old, roughly 65% of males are married; while roughly 76% of women are married, reflecting an earlier average age of marriage and child birth for women. Astoundingly, in this age group over 90% of males participate in the labor force, while roughly 30% of women participate. These reflect prime working age in the country as both values are significantly higher than for the population at large. Marriage and large families are common. Roughly three-quarters of women have married by age 30, and for older cohorts this rises above 90%. While the total fertility rate in Jordan has declined since the 1980s when it exceeded 7, it remains high enough for population growth at 3.3 today (WB, 2018).

The 2004 Jordanian Census recorded educational outcomes for those who have completed their education as a categorical variable divided into bins. Figure 12 depicts educational attainment for individuals born in birth cohorts ranging from 1967 to 1980.[45] As can be seen from the upper right and lower left panels of the figure, a visible discontinuity occurs from the 1973 to 1974 birth cohorts. It is this forced jump in education attainment we exploit to derive causal estimates of educational attainment. It is also clear that the transition from Preparatory to Basic education did not completely transition from one cohort

[44] Corresponding summary statistics for the EUS sample are available upon request.
[45] We present these figures for a wide band which we use in some robustness checks and which we believe provides useful context on the pre and post reform trends in educational attainment in the country. By the nature of our empirical approach, we typically examine narrower intervals around the educational reform.

to the next. It is for this reason that we adopt a fuzzy regression discontinuity designs strategy. We consider a number of explanations for the imperfect adjustment ranging from incomplete compliance with the compulsory schooling rule, to insufficient resources to implement the reform, to survey response error in Section 5.1.

Figure 12 presents mean years of education for individuals in our sample frame who have completed any degree. There are no visibly large discontinuities in the share of the population attaining secondary or higher education after the reform as well; suggesting the increase in education we study is predominantly confined to the impact of increasing the duration of compulsory years from 9 to 10. As discussed in Section 2, at the same time that the reform increased the duration of basic compulsory education, it reduced the duration of secondary school by 1 year, meaning that all higher degrees such as secondary, masters, and bachelor's degrees still represented the same number of years of educational attainment. Panel B of Table 1 contains summary statistics on educational attainment for our sample.

The two samples generally paint a similar picture of the Jordanian population, with a few nuances worth pointing out. First, the EUS waves begin in 2006 and run through 2016, so they observe our birth cohorts of interest across a series of later points in their lives. Second, the EUS surveys contain additional information on wages, which we discuss and examine in more detail in Section 5.3. A third interesting difference is that the EUS sample suggests 41% of males and 56% of women work in the public sector. In the census the corresponding numbers are 35 and 30%. Some of this difference may be attributable to how the questions are worded, and government statistics suggest the EUS numbers are closer to reality, with women disproportionately seeking the public sector over the private.

Anecdotally, public sector employment is often considered more socially acceptable for women.

To assess the availability of educational resources surrounding the reform, we compile a number of statistics from disparate sources. First, using data on the date of construction for the corpus of schools in Jordan from Ministry of Education's 2010 Census of Schools in Jordan, we retrospectively construct the number of schools present in each district in each year following the general approach of Assaad and Saleh (2016). From annual statistical yearbooks, we complement data on the number of schools with information on the total number of teachers and pupils in the country. This data allows us to construct average values for the number of teachers per school, students per school, and the teacher/pupil ratio at a regional level over time – a set of proxies for the evolution of educational resources and demands on the educational system.

2.4 Empirical Strategy

2.4.1 Methodology

As discussed in Section 2, the educational reform extended de jure years of compulsory schooling in Jordan from nine to ten, which theoretically should create a discontinuity in years of education attained across cohorts. In practice, de facto enforcement of the new Basic, 10-year, compulsory schooling was imperfect, with some individuals still only attaining nine years of schooling. Nevertheless, compliance was still significant enough to generate the discontinuity visible in Figures 12 and 13, and it is this development that we exploit to examine the causal impact of educational on labor market outcomes.

From an econometric standpoint, these nuances of the data means that the treatment, T_i – an additional year of completed required schooling – is determined only in

part by the value of a forcing variable, x_i, cohort of birth, since when an individual was born determines whether they were exposed to the schooling reform. Even with imperfect compliance, it is possible to study the causal impact of education reform using regression discontinuity approaches as long as the probability of treatment is discontinuous at a cutoff point, denoted by x_0. As can be seen in Figure 12, there is a discrete break in educational attainment across the 1973 and 1974 birth cohorts in which roughly half of the typical population attaining preparatory suddenly began attaining basic instead. And as discussed previously, the fraction of the population attaining other degrees exhibits no large discontinuities around this time threshold.

We thus employ a fuzzy regression discontinuity designs to evaluate the impact of increased educational attainment resulting from the 1988/1989 school reforms.[46] Following Imbens and Lemieux (2008), we assign $T_i = 1$ when individual i receives an additional year of schooling and $T_i = 0$ when they do not. Probability of assignment to the treatment is then given by:

$$\Pr(T_i = 1|x_i) = \mathrm{E}(T_i|x_i) \tag{1}$$

and the following assumption is satisfied:

$$\lim_{x_i \to x_0^+} \mathrm{E}(T_i|x_i) \neq \lim_{x_i \to x_0^-} \mathrm{E}(T_i|x_i) \tag{2}$$

when the probability of treatment is discontinuous at the threshold x_0. We further denote y_{0i} to be the outcome if individual i has not been treated and y_{1i} be the treated outcome for individual i. Given y_i as an observed outcome, it can be expressed as follows:

[46] Regression discontinuity designs were first introduced by Thistlethwaite and Campbell (1960), while Hahn et al., (2001) contributed the key theoretical framework and concepts for the application of fuzzy RD designs.

$$y_i = \alpha_i + \rho_i T_i \tag{3}$$

where $\alpha_i = y_{0i}$ and $\rho_i = y_{1i} - y_{0i}$. In this setup, ρ_i is the treatment effect for individual i. However, in practice, we never observe y_{1i} and y_{0i} together. Instead, we can estimate the average effect of the treatment and derive the treatment effect ρ over our sub-population of interest.

An essential property for clean identification is that the conditional mean function $\mathrm{E}(\alpha_i|x_i)$ is continuous at $x_i = x_0$. Stated differently, our fuzzy RD designs approach requires that in the absence of the reform, average outcomes would otherwise be similar for individuals just below the cutoff in comparison to those just above the cutoff. As long as this assumption holds, then it can be shown that

$$\lim_{x_i \to x_0^+} \mathrm{E}(y_i|x_i) - \lim_{x_i \to x_0^-} \mathrm{E}(y_i|x_i)$$

$$= \rho \left(\lim_{x_i \to x_0^+} \mathrm{E}(T_i|x_i) - \lim_{x_i \to x_0^-} \mathrm{E}(T_i|x_i) \right) \tag{4}$$

$$\Rightarrow \quad \rho = \frac{\lim_{x_i \to x_0^+} \mathrm{E}(y_i|x_i) - \lim_{x_i \to x_0^-} \mathrm{E}(y_i|x_i)}{\lim_{x_i \to x_0^+} \mathrm{E}(T_i|x_i) - \lim_{x_i \to x_0^-} \mathrm{E}(T_i|x_i)} \tag{5}$$

where ρ provides an estimate of the local average treatment effect. In analysis, we revisit this assumption and discuss its plausibility in the context of this particular Jordanian educational reform. We also present several robustness checks along this dimension.

The estimation of the average treatment effect ρ is typically a nonparametric regression problem. However, Imbens and Lemieux (2008) show that bias for the simple kernel estimator when one non-parametrically estimates a regression function in the interior of the support is relatively high and is related to the bandwidth h. One practical solution to

this issue is a local linear regression (Fan and Gijbels, 1996; Imbens and Lemieux, 2008). In a fuzzy RD designs, using a local linear regression with a uniform kernel is comparable, under certain conditions, to obtaining the treatment effect via a two-stage least squares (2SLS) approach.[47] This equivalency also holds when we include additional covariates as discussed in the subsequent analysis (Hahn et al. 2001; Imbens and Lemieux, 2008; Calonico, et al., 2018).

While the theoretical treatment above describes the case of fuzzy regression discontinuity designs for an indicator treatment variable for simplicity, it also holds in a continuous setting. The existing literature on the evaluation of compulsory schooling changes have widely used a measure of treatment intensity, namely years of education as their measure of treatment assignment (Albouy and Lequien, 2009; Van Kippersluis et al., 2011; Ali and Gurmu, 2016). We follow this approach as it has the advantage of being directly comparable with pre-existing research and because using variable intensity allows us to more easily interpret the marginal effect of schooling on labor market outcomes.

In our 2SLS approach, we use an indicator for the introduction of the 10 year compulsory schooling reform, $Basic_{Reform_i} = 1(x_i > x_0)$, and the first order polynomial interaction of the reform with birth cohorts as a set of instruments. The first stage equation is given,

$$T_i = \alpha + \beta_1 Basic_{Reform_i} + \beta_2 x_i + \beta_3 Basic_{Reform_i} * x_i \qquad (6)$$

$$+\boldsymbol{Z}_{dt}\boldsymbol{\Gamma} + \boldsymbol{\Omega_d} + \varepsilon_{it}$$

[47] Formally, it works when the same bandwidth is employed for the estimation of the conditional expectation of the outcome variable and for the conditional expectation of the treatment variable on both sides of the threshold, which is the case in our analysis.

and, similarly, the reduced form equation is

$$Y_i = \theta + \pi_1 Basic_{Reform_i} + \pi_2 x_i + \pi_3 Basic_{Reform_i} * x_i \quad (7)$$

$$+\boldsymbol{Z}_{dt}\boldsymbol{\Gamma} + \boldsymbol{\Omega_d} + \omega_{it}$$

where i indexes individuals.[48] Y_i is our labor market outcomes of interest, which we vary by specification - labor force participation, employment status, an indicator for self-employment, and hours worked per week.[49] Zdt represents a vector of time variant district and governate specific controls, which we include to help isolate the impact of any correlated investments in the educational system from the impact of the additional year of schooling. Ωd is a vector of district fixed effects, and where d indexes districts.

Omitted factors are a concern when they are (a) heterogeneous across districts, (b) time-variant, and (c) potentially correlated with the implementation of the compulsory schooling component of the reform. Other educational investments occurring concurrently with the compulsory schooling extension are a first order candidate. As detailed in Section 3, we attempt to mitigate this issue with the inclusion of fixed effects and by controlling for the per capita stock of same gender or co-ed schools in the district, as well as governate level student-teacher and class unit-teacher ratios.

2.4.2 Bandwidth selection

The selection of bandwidth in regression continuity designs presents a tradeoff between bias and variance. A wider bandwidth may have a larger bias because more observations are far from the cutoff, but it will have lower variance because it can employ a

[48] In this framework, the estimated treatment effect is given by $\rho = \frac{\pi_1}{\beta_1}$.

[49] All outcomes observed at the time of survey enumeration: 2004 for the Census; various years for the EUS. Further details can be found in the table notes.

larger number of observations. We take two approaches for selecting an appropriate bandwidth. First, we use a cross-validation procedure based on mean-square error to let the data suggest a bandwidth for each outcome variable (Calonico, 2016). Across the outcomes, this technique produces bandwidths which range from 4 to 7 years.

As a second approach, we consider narrower ranges of birth cohorts on a theoretically basis. One issue with allowing for a comparison across more years is that individuals become less and less comparable with time. Factors such as cohort effects and macroeconomic conditions at the time of entry in the labor market have more scope to vary. Perhaps most importantly, other impacts of the Education Act No. 27 in 1988 beyond the compulsory schooling law itself would, over longer time horizons, have had time to have an impact. This could include the reformed structure and content of secondary school as well as any additional educational investments over the period such as school building, teacher training or hiring.

Several things are worth noting in this regard. First, anecdotally, the most commonly documented impact of the reform was the extension of compulsory schooling and the curricula/structural reforms in secondary school. Second, there were no other major educational reforms in the years immediately preceding or following the 1988/1989 reforms. The closest reform occurred in 1994, a year during which a new unified educational philosophy was issued. Using our preferred bandwidth, 4 years, excludes individuals exposed to this reform, and even using a longer bandwidth of 7 years would still only include individuals exposed to the new reform in the final 2 of their 10 years of education. In the analysis, we present results for 4 years as the bandwidth, and then run robustness checks using a range from 3 to 7 years. In all analysis, we employ the same bandwidth for the

outcome and treatment regressions and the same bandwidth length on either side of the reform.

2.4.3 A graphical exploration

As a first pass, it is useful to examine how years of education vary over time in relation to the educational reform. Figure 13 plots years of education across Jordanian birth cohorts using the 2004 Census. A discontinuity in the average years of education is visible between those born in 1973 and 1974.[50] There is a discernable increase in years of education completed by individuals born after 1973 who attended 10 years compulsory schooling (those affected by the 1988/1989 compulsory schooling reform) compared to individuals born in 1973 and earlier years. This increase is on the order of about one-fifth of a year of education for the entire cohort.

This jump in years of education is consistent with the evidence we have previously presented in two ways. First, the educational attainment of roughly 70% of the population would not be expected to change by the reform. This is because about 5% of the Jordanian population either never attended or completed any formal schooling (and were thus not impacted), while about 10% completed no more than primary before and after the reform, and about 50% completed secondary degrees or higher (recall from the previous discussion that the total years of attainment were unaffected for this group as well). The only affected group should have been those stopping education at the preparatory level, a group

[50] The dots in Figure 2 represent the unconditional mean of years of education for each birth cohort group and the solid lines represent fitted quadratic regression lines with confidence intervals.

comprising about 30% of the population. And within this population, only about half were impacted by the Basic reform.[51]

2.4.4 First-stage estimates

Table 7 presents the first-stage OLS estimates for the sample of male, female, and both. It indicates that the 1988/1989 compulsory schooling reform increased the average years of education of individuals by between two and three months depending on whether the cohort trends are specified as linear or quadratic. Even though the coefficients on reform indicator are similar and significant in both models, but the linear model has the lowest value of AIC.[52]

Based on AIC, we choose to present results from the linear model as the baseline model and simply run alternative specifications of the cohort trends as robustness check. For the separate samples of male and female, coefficients on reform also have similar magnitudes for both linear and quadratic models, and again, the linear regressions still produce lower AIC value than the quadratic models. Those coefficients indicate that the 1988 reform increased the average years of schooling of females and males by about one to two months as well as three to four months respectively.

[51] In the diagram and the analysis, we exclude individuals with less than primary educational attainment. We do this primarily because estimates of the return to an additional year of school for these individuals, many of whom are illiterate, are unlikely to be representative of that for the reform in question which expanded schooling only among higher attainment levels. We explore alternative subpopulations of educational attainment groups as robustness checks in the next section.

[52] The Akaike information criterion (AIC) is a test of model selection to balance the tradeoff between goodness of fit and simplicity of the model. It equals to twice of the difference between the number of estimated parameters and the value of log of maximum likelihood function at the optimal. The optimal model is the one with the lowest value of AIC.

2.5 Results

2.5.1 Schooling and Engagement with the Labor Market

Table 8 presents the results from the fuzzy regression discontinuity designs specification described in equations 6 and 7, using a 4 year window on either side of the 1988/1989 reform (including only birth cohorts spanning 1970 to 1977), and employing a liner functional form with interactions as discussed in Section 4.1. Each row and column represent an individual regression, all of which contain district fixed effects. The first four columns provide estimates of labor market status in a sample of individuals from the 2004 Census, while the remaining columns use a pooled sample drawn from the closest EUS waves temporally, 2006 and 2007.[53] These regressions additionally control for a flexible set of polynomials in age to account for the fact that individuals are observed at different times since graduation in each wave.

Panel A presents estimates for both men and women together, while Panel B and C isolate the sample of men and women respectively. Columns (1) and (5) present estimates of the impact of an additional year of compulsory schooling on labor market participation. Looking first at the full sample, the coefficient of an additional year of education on labor force participation is estimated to be positive and significant at 0.032, suggesting that each additional year of schooling is associated with a 3.2% point increase in the likelihood an individual participates in the labor force.

Given a mean labor force participation rate of 64.1%, each additional year of schooling would then be associated with a roughly 5% increase in labor force participation. In practice, because compliance with the actual reform was limited, these estimates imply

[53] We discuss the reasoning behind this approach in Section 5.2.

that in the first year of its implementation, the roughly 0.2 year increase in schooling would have increased labor force participation by about 0.6 percentage points (a 1% gain). The results for employment status in columns (2) and (6) generally mirror those for labor force participation as well.

As can be seen from Panels B and C, these impacts are driven predominantly by women entering the labor force and being more likely to be employed with the additional education. This is both unsurprising and surprising. It is unsurprising because nearly all males participate in the labor market, so there is not a large scope for additional education to increase this. Incredibly, in the age groups in our regression sample, over 95% of males engage in the labor market and between 85 and 90% work.[54] At the same time, it is surprising to observe because social norms operate to constrain women's opportunities for economic participation in Jordan.

Columns (3) and (7) present the impact of additional schooling on self-employment. An extra year of education is associated with a small increase in self-employment, largely driven by males. Similarly, columns (4) and (8) demonstrate an increase in hours worked resulting from an additional year of schooling, driven entirely by males. That we observe such a large impact on the intensive margin rather than the extensive margin for males makes intuitive sense given that nearly all males already participate in the labor force.

Contrasting the two samples produces additional insights. While the Census and the EUS survey estimates are relatively close for the overall labor force impacts and for females, the EUS yields generally larger statistics for the impact of education on male labor force

[54] The estimate for male labor force participation is one location where the Census and EUS results diverge slightly, with the impact on LFP of males larger in the EUS and close to zero in Census.

participation, employment status, and hours worked respectively. One consideration is that the EUS provides a snapshot of individuals slightly later in life, and we provide hypotheses and explore this feature of the data in more detail in the following section.

2.5.2 Education and Labor Market Outcomes over the Life Cycle

The 2004 Census captures individuals on either side of the 1988/1989, meaning we observe individuals with an average 30 years of age, plus or minus the bandwidth we select. A novel feature of using the Employment and Unemployment Surveys to also examine this question is that we have snapshots of the same cohorts of individuals at later periods of their life. Patterns of labor market outcomes over the life cycle are very different for men and women over time in Jordan. These are depicted in Figures 14 and 15 respectively across cross-sections of the population observed in four of the EUS waves.

We focus first on Figure 14. By age 20, roughly 70% of males are in the labor force. Nearly all males are working by 23, and this rate plateaus for men until about age 35. At which point, it begins to decline slowly over the life-cycle. By age 40 it has fallen about 5% points and by age 50 it has declined to a level between 60 and 70%.

This pattern is remarkably different than that for women. As can be seen in Figure 15, women enter the labor force at roughly the same age as men, although only 30 to 40% of them enter at this point (in a promising trend, the highest levels of female labor force entrance are observed in the most recent EUS waves). This is where the similarities end. Female labor force participation peaks at 23, and unlike for men where it plateaus, it immediately begins declining at this point. By age 30, about a quarter of those who entered have left, and by age 40 nearly half have. By age 50, labor force participation rates approach 10%.

These patterns likely could reflect the influence of market factors including higher rates of unemployment for women and lower wages, although these differences are not especially large while the different life cycle patterns are striking. Indeed, the stark patterns are probably more consistent with a set of strong social pressures for women to marry, have children, and provide non-market labor within the household which raise the opportunity cost of remaining in the labor force as women age (which may also manifest in a perceived lack of acceptability of engaging in the labor market).

To examine this further, we utilize the EUS in yet another way in Table 9. Here we run our fuzzy RD designs to examine the causal impact of an additional year of schooling from the 1988/1989 reforms in the first two available waves of the EUS 2006 and 2007 and contrast that with those form the 2014 and 2016 EUS waves. The first five columns reproduce those from Table 8, and we observe patterns similar to the Census where extensive margin labor force and employment gains are large for women from the schooling reform. Gains for men exist but are smaller, and we actually see large intensive margin gains for men in terms of hours worked. This is consistent with the patterns we observed in Figures 14 and 15. Men and women are both induced to engage more heavily with the market at this earlier stage of their life (around age 32), just along different margins.

When we turn to the last set of columns, the labor force participation gains for women are essentially estimated to be similar as in Census, while those for men are more than four times larger. The same is true for employment. This suggests that the additional schooling may have helped men remain in the labor market.

2.6 Discussion and Robustness

2.6.1 Functional Form of the RD Specification

In this section we explore the robustness of our empirical specification to alternative econometric modeling choices. Section 4.4 discussed the first stage results of the impact of the change to compulsory schooling on years of educational attainment. In our analysis, we used a parametric model with a linear polynomial functional form with treatment interaction terms -- on the basis that it produced lower AIC values than those for the quadratic models. In Table 10, we replicate these baseline estimates in Panel B following this approach. Panels A, C, and D, present the results if we instead use a linear, quadratic, or quadratic interaction approach respectively. With the exception of small changes in the estimated impact on wages, the results are qualitatively and quantitatively similar regardless of specification, suggesting that the choice of functional form for the regression discontinuity is not a first order concern in this setting.

A second econometric concern is that a number of our labor market outcomes of interest are constructed as binary variables – including labor force participation, employment status, and our indicator for self-employment. A typically raised issue when conducting analysis on a binary dependent variable is that if this structure is not directly modeled, regression estimates are capable of generating predictions outside the range of 0 to 1. As a robustness check, IV Probiit specification reproduces the results of Table 8. Reassuringly, in all cases, these results are remarkably similar to those produced using our baseline specification.

2.6.2 Alternative Definitions of the Treatment Population

Arguably, the creation of Basic education in 1988 should have had by far the largest impact within the population of individuals who were previously electing to stop school at the end of preparatory education (9 years). In theory, individuals who were stopping school earlier, such as at the end of the 6 year window (primary) or who had preprimary or lower educational attainment could have been induced to stay in school longer by the reform act.[55] This would be the case if, for example, the reform been perceived to increase educational quality and individuals expected this to translate into better labor market outcomes.

In practice, this change seems unlikely in the case of 1988 Jordan for a number of reasons. First, slow implementation of the new Basic program suggests that many schools may not even have immediately known or been capable of enacting the reform. This suggests that most students and parents would likely have been even worse informed of the reform and its repercussions. Second, the share of the population leaving school before reaching 9 years of education exhibited no visible discontinuity. If individuals were induced to stay in school longer, we might expect to see a more rapid increase in years of educational attainment. It is still possible however that the addition of another year of school in the regular curriculum may have increased earlier dropout, as the time cost associated with obtaining the regular school age stopping degree had been increased.[56]

---

[55] It is also possible that the introduction of the 10th year of compulsory education impacted the decision to attend secondary school or to continue in the educational system beyond ten years into a master's or bachelor's degree in the university system. Although plausible, rigidities in the Jordanian educational system limit the scope for individual and household responses to this change because entrance into secondary school tracks are based on relative performance in the classroom, and acceptance into public universities are based on a competitive exam.

[56] While it is possible that both forces could have operated and offset one another, the empirical strategy in Section 4 should be relatively robust to this case. This is because the reform was implemented shortly after being passed, and thus there was not a large period of time in which households would have been able to forecast this change. This means that individuals in the sample only learned of the change to compulsory

When we include all individuals, even those who are illiterate or do not complete primary school. In all cases, coefficient magnitudes decrease slightly, yet remain significant and quite similar to those from our main specification in Table 8. This would be as expected since the addition of a larger "untreated" population into the analysis should introduce noise and attenuate the coefficients.

2.6.3 Robustness to Bandwidth Selection

As discussed in Section 4.2, the selection of bandwidth presents a tradeoff between bias and variance. We elected to present results for a 4 year bandwidth following intuition and both empirical selection procedures. Nevertheless, should unobserved educational investments not captured by our time variant measures of schools per capita and teachers per capita have occurred in the sample period, in longer bandwidth specifications, these could bias our coefficients in either direction depending on whether these investments were positively or negatively correlated with compliance with the new basic requirements.[57] We present the primary results for bandwidths of 3, 4, 5, and 7 years in upon request. Fortunately, the primary results are not incredibly sensitive to our choice of bandwidth.

2.7 Conclusion

This paper studies the causal impact of an educational reform in Jordan during the 1988 to 1989 school year which extended years of compulsory schooling from nine to ten. Our analysis suggests that additional schooling in the late 1980s induced more men and women to enter the labor force participation, raised their chances of finding employment, and boosted wages. Gains were larger in the intensive margin for men, who realized increased

schooling in the final few years of their preparatory education, at which time many of the early dropouts would have already left school or decided not to enroll in the first place.

[57] Furthermore, the estimates we produce should then be interpreted as the net impact of these other overall educational reforms enacted in the period, not of schooling alone.

in hours worked as a result. These patterns provide some evidence that educational improvements can still translate into stronger labor markets, but also highlight the strength of traditional gender norms in Jordanian society. Institutions and norms sometimes evolve slowly, and the broader impact of even temporary improvements in labor market participation may be larger than observed today, such as influences on future generations. To this extent, future research may explore factors which affect the extent to which these norms and institutional forces dictate the health of the Jordanian labor market and prospects for female employment.

Table 6. Census Summary Statistics

	Full Sample			Men			Women		
Panel A: Individual Characteristics	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Age	101,728	30.06	4.00	51,441	30.03	4.01	50,287	30.08	3.99
Married	101,728	0.70	0.46	51,441	0.65	0.48	50,287	0.76	0.42
Number of children	101,728	1.87	1.99	51,441	1.43	1.74	50,287	2.33	2.12
Age at first birth of child	60,729	24.05	4.17	26,455	26.33	3.48	34,274	22.30	3.80
Labor force participation	101,728	0.62	0.49	51,441	0.93	0.26	50,287	0.31	0.46
Employed	101,728	0.49	0.50	51,441	0.80	0.40	50,287	0.17	0.37
Self employed[a]	63,155	0.09	0.29	47,750	0.12	0.32	15,405	0.01	0.12
Public sector employee[a]	63,155	0.33	0.47	47,750	0.35	0.48	15,405	0.30	0.46
Hours worked[a]	48,512	48.91	13.95	40,167	50.14	14.21	8,345	42.98	10.79
Panel B: Educational Characteristics									
Mean years of schooling[b]	94,460	11.56	2.98	47,719	11.40	3.04	46,741	11.72	3
Educational Attainment									
Primary (6 years)	101,728	0.08	0.27	51,441	0.09	0.29	50,287	0.07	0.26
Preparatory (9 years)	101,728	0.20	0.40	51,441	0.22	0.41	50,287	0.19	0.39
Basic (10 years)	101,728	0.10	0.30	51,441	0.12	0.32	50,287	0.09	0.29
Old Secondary (12 years)	101,728	0.11	0.32	51,441	0.12	0.32	50,287	0.11	0.32
New Secondary (12 years)	101,728	0.14	0.34	51,441	0.14	0.34	50,287	0.14	0.34
Bachelors or More	101,728	0.29	0.45	51,441	0.25	0.44	50,287	0.32	0.47
Panel C: Household Characteristics									
Family size	101,728	5.80	2.71	51,441	5.66	2.79	50,287	5.94	2.61
Own home	101,728	0.78	0.41	51,441	0.78	0.41	50,287	0.78	0.42
Polygamous household	101,728	0.01	0.10	51,441	0.01	0.09	50,287	0.01	0.12
Urban residence	101,728	0.78	0.41	51,441	0.78	0.41	50,287	0.78	0.42

Notes: Full sample of individuals borth between 1967 and 1980. Sample excludes the illiterate, individuals with an indeterminate age, current students, and individuals who migrated into Jordan after school starting age (age 6). Mean years of schooling imputed using educational attainment, time to degree based on birth cohort, and length of degree program in years. (a) Percentages calculated among those in the labor force. (b) Among those with at least primary attainment.Source: Author's calculations using 2004 Jordanian Census (IPUMS).

Table 7. First Stage - The Impact of the Compulsory Schooling Reform on Years of Education

	Linear			**Quadratic**		
	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample	Men	Women	Full Sample	Men	Women
Reform Indicator	0.192***	0.114	0.270***	0.261**	0.168	0.353**
	(0.053)	(0.076)	(0.072)	(0.105)	(0.153)	(0.145)
Birth Cohort	-0.015	-0.007	-0.024	-0.084	-0.044	-0.120
	(0.016)	(0.023)	(0.022)	(0.091)	(0.132)	(0.126)
Birth Cohort Squared				-0.014	-0.007	-0.019
				(0.018)	(0.026)	(0.025)
Reform * Cohort	0.081***	0.060*	0.107***	0.146	0.048	0.241*
	(0.022)	(0.032)	(0.031)	(0.106)	(0.154)	(0.146)
Reform * Cohort Squared				0.015	0.024	0.006
				(0.025)	(0.036)	(0.034)
Number of observations	54,056	27,285	26,771	54,056	27,285	26,771
AIC	268374	136914	131267	268378	136918	131270

Notes: Sample as described in Table 1, restricted to those with at least 6 years of completed education and born between 1970 and 1977.
Source: Author's calculations using 2004 Jordanian Census (IPUMS).
*** p<0.01, ** p<0.05, * p<0.1

Table 8. The Effects of the 1988/1989 Jordanian Educational Reform on Labor Market Outcomes

	Census				Employment and Unemployment Survey				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Full Sample	**LFP**	**Employed**	**Self Employed**	**Hours Worked**	**LFP**	**Employed**	**Self Employed**	**Hours Worked**	**Wages**
Years of Education	0.032***	0.032***	0.012***	0.641***	0.034***	0.029***	0.013***	3.689***	44.852***
	(0.003)	(0.003)	(0.001)	(0.107)	(0.006)	(0.006)	(0.002)	(0.314)	(2.445)
Mean of Dep Var.	0.641	0.521	0.060	48.92	0.571	0.526	0.033	44.78	231.8
Number of Observations	54,056	54,056	54,056	27,622	44,777	44,777	44,777	23,534	23,437
Panel B: Males									
Years of Education	0.003*	0.004	0.024***	1.562***	0.031***	0.036***	0.032***	6.755***	61.405***
	(0.001)	(0.003)	(0.003)	(0.137)	(0.004)	(0.006)	(0.005)	(0.532)	(3.812)
Mean of Dep Var.	0.967	0.857	0.114	50.12	0.950	0.895	0.064	46.17	234.9
Number of Observations	27,285	27,285	27,285	22,959	21,746	21,746	21,746	19,473	19,412
Panel C: Females									
Years of Education	0.080***	0.077***	0.004***	-0.207	0.073***	0.061***	0.003***	-0.106	36.481***
	(0.003)	(0.003)	(0.001)	(0.358)	(0.005)	(0.005)	(0.001)	(0.571)	(3.991)
Mean of Dep Var.	0.308	0.178	0.005	43.02	0.214	0.176	0.005	38.09	216.7
Number of Observations	26,771	26,771	26,771	4,663	23,031	23,031	23,031	4,061	4,025
Age and Age Squared	NA	NA	NA	NA	Yes	Yes	Yes	Yes	Yes
Birth Cohort Polynomials	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Fuzzy RD designs (2SLS regression) estimates including linear interaction terms. Sample as described in Table 1, restricted to those with at least 6 years of completed education and born between 1970 and 1977.

Source: Author's calculations using 2004 Jordanian Census (IPUMS) and the 2006 and 2007 EUS Surveys (ERF).
*** p<0.01, ** p<0.05, * p<0.1

Table 9. Short vs. Long-Term Effects of Educational Attainment on Labor Market Outcomes in Jordan

	EUS 2006 and 2007					EUS 2014 and 2016				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Full Sample	**LFP**	**Employed**	**Self Employed**	**Hours Worked**	**Wages**	**LFP**	**Employed**	**Self Employed**	**Hours Worked**	**Wages**
Years of Education	0.039***	0.035***	0.014***	3.276***	41.567***	0.076***	0.076***	-0.000	-3.754***	70.439***
	(0.006)	(0.006)	(0.002)	(0.310)	(2.443)	(0.009)	(0.009)	(0.004)	(0.341)	(5.719)
Panel B: Males										
Years of Education	0.026***	0.032***	0.031***	5.857***	55.850***	0.121***	0.129***	-0.001	-0.922***	69.619***
	(0.004)	(0.006)	(0.005)	(0.505)	(3.727)	(0.012)	(0.013)	(0.008)	(0.330)	(7.305)
Panel C: Females										
Years of Education	0.078***	0.068***	0.003***	0.646	45.171***	0.079***	0.074***	0.002*	2.356***	21.502**
	(0.006)	(0.005)	(0.001)	(0.641)	(5.190)	(0.008)	(0.008)	(0.001)	(0.867)	(10.737)
Age and Age Squared	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Cohort Polynomials	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Fuzzy RD designs (2SLS regression) estimates including linear interaction terms. Sample as described in Table 1, restricted to those with at least 6 years of completed education and born between 1971 and 1976.

Source: Author's calculations using the 2006, 2007, 2014, and 2016 EUS Surveys (ERF).
*** p<0.01, ** p<0.05, * p<0.1

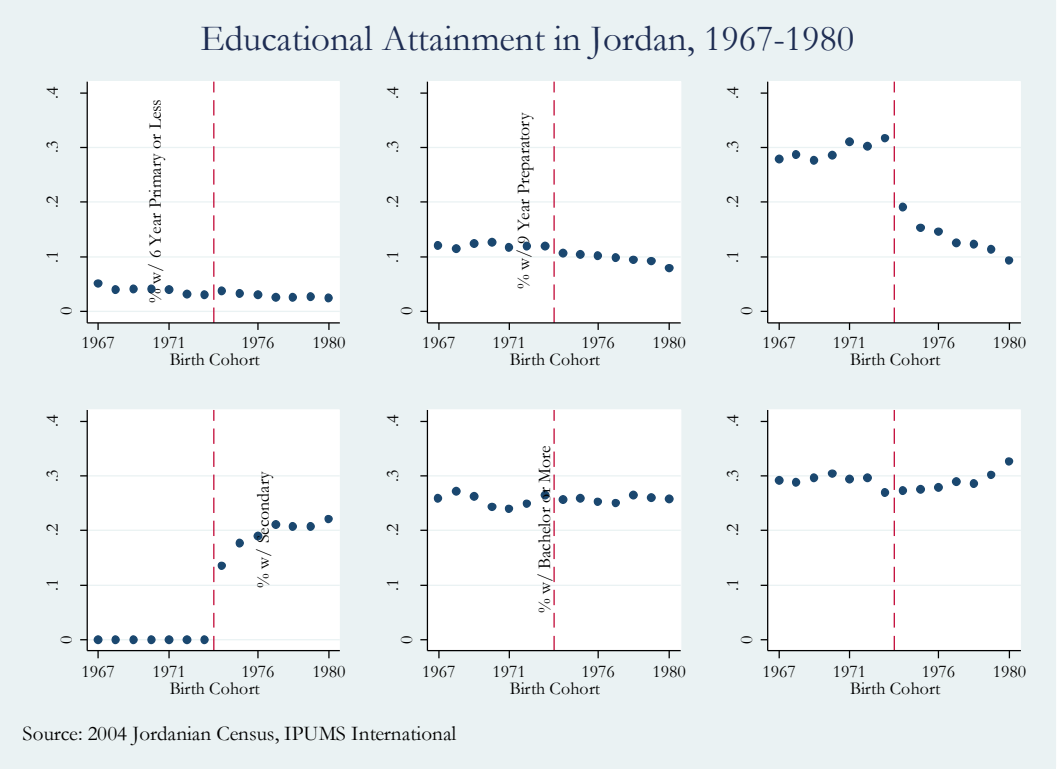
Table 10. Sensitivity to Functional Forms of the Regression Discontinuity Design

	Census				Employment and Unemployment Survey				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Linear	**LFP**	**Employed**	**Self Employed**	**Hours Worked**	**LFP**	**Employed**	**Self Employed**	**Hours Worked**	**Wages**
Years of Education	0.031***	0.032***	0.013***	0.648***	0.035***	0.029***	0.013***	3.720***	45.037***
	(0.003)	(0.003)	(0.001)	(0.107)	(0.006)	(0.006)	(0.002)	(0.315)	(2.455)
Panel B: Linear Interaction									
Years of Education	0.032***	0.032***	0.012***	0.641***	0.034***	0.029***	0.013***	3.689***	44.852***
	(0.003)	(0.003)	(0.001)	(0.107)	(0.006)	(0.006)	(0.002)	(0.314)	(2.445)
Panel C: Quadratic									
Years of Education	0.032***	0.032***	0.013***	0.642***	0.035***	0.029***	0.013***	3.687***	44.860***
	(0.003)	(0.003)	(0.001)	(0.107)	(0.006)	(0.006)	(0.002)	(0.314)	(2.446)
Panel C: Quadratic Interaction									
Years of Education	0.032***	0.032***	0.012***	0.641***	0.035***	0.030***	0.013***	3.692***	44.828***
	(0.003)	(0.003)	(0.001)	(0.107)	(0.006)	(0.006)	(0.002)	(0.314)	(2.444)
Age and Age Squared	NA	NA	NA	NA	Yes	Yes	Yes	Yes	Yes
Birth Cohort Polynomials	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Fuzzy RD designs (2SLS regression) estimates including linear interaction terms. Sample as described in Table 1, restricted to those with at least 6 years of completed education and born between 1970 and 1977.

Source: Author's calculations using 2004 Jordanian Census (IPUMS) and the 2006 and 2007 EUS Surveys (ERF).
*** p<0.01, ** p<0.05, * p<0.1

Figure 12. Educational attainment in Jordan, 1967 - 1980



Source: 2004 Jordanian Census, IPUMS International

Figure 13. The impact of educational reform in 1988/1989

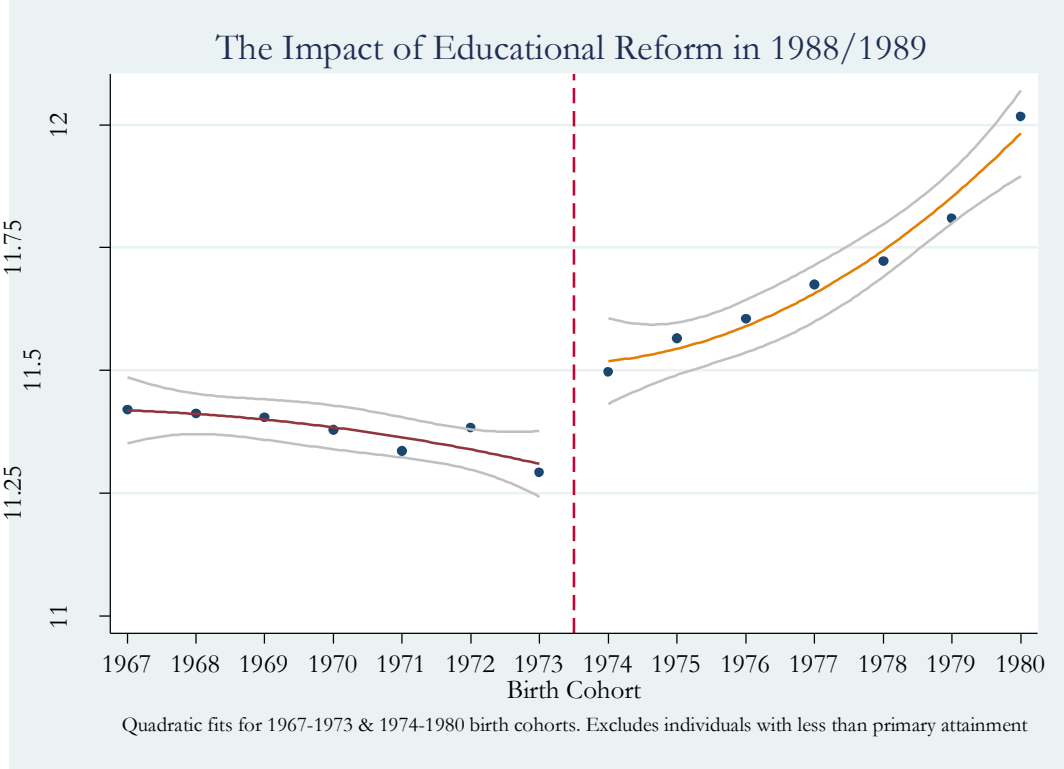


Figure 14. Male labor market engagement over the life cycle

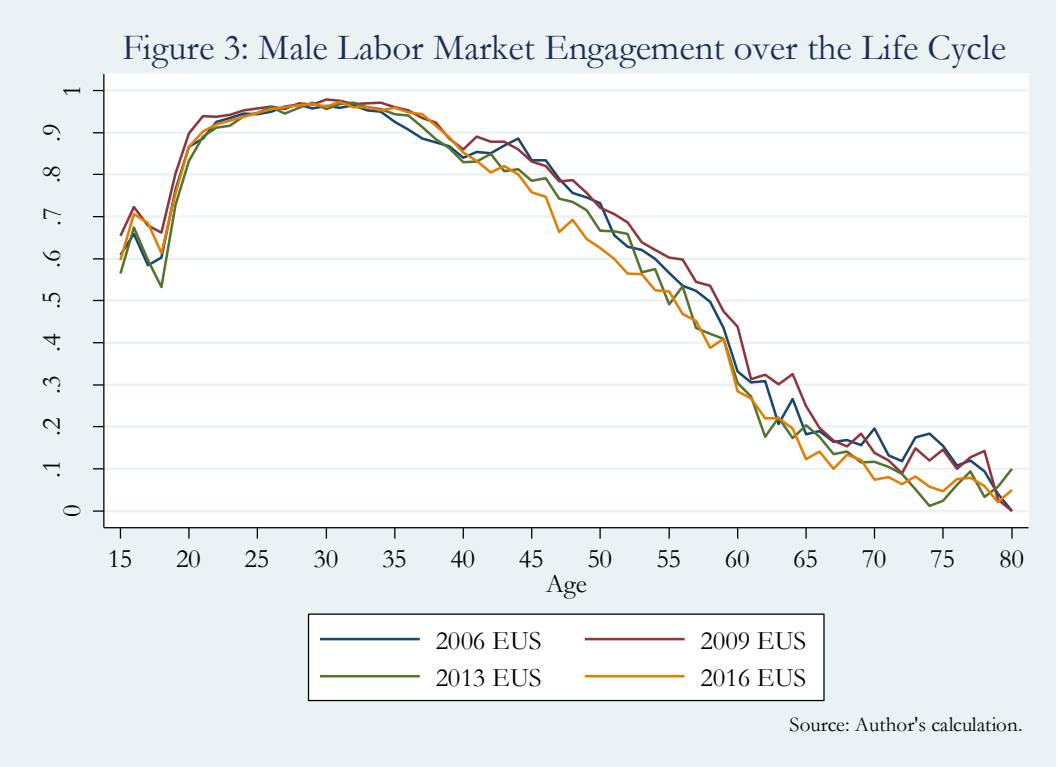
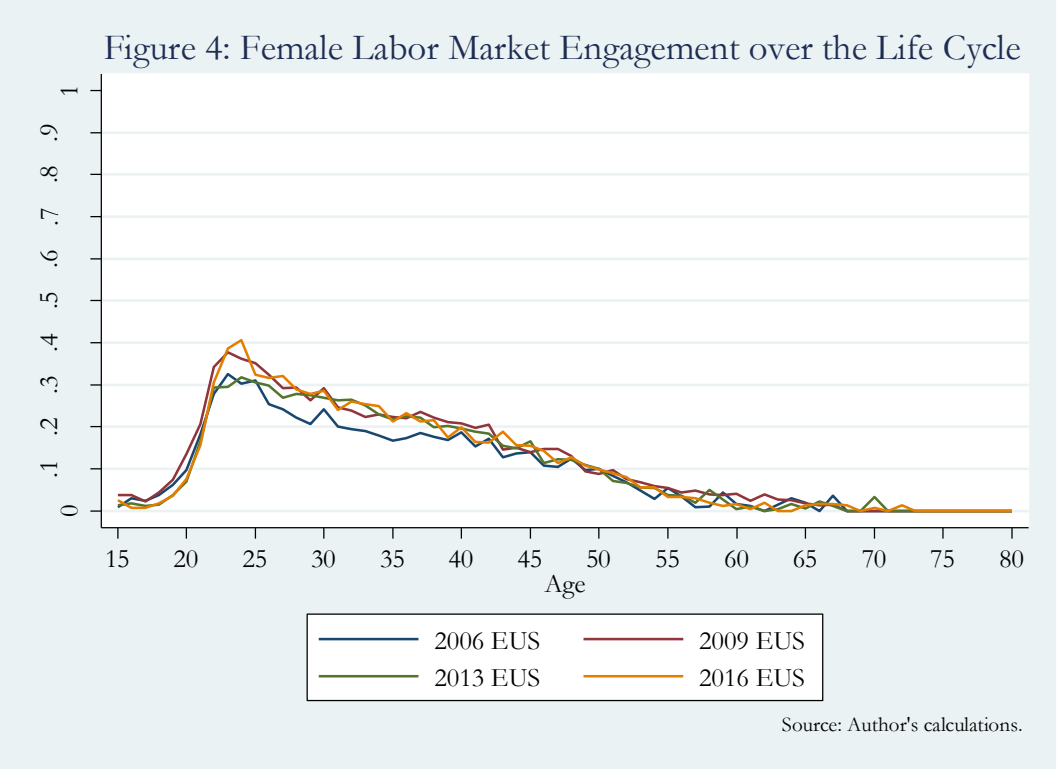


Figure 15. Female labor market engagement over the life cycle



References

Albouy, Valerie, and Laurent Lequien (2009). "Does compulsory education lower mortality?" *Journal of Health Economics* 28(1): 155-168.

Ali, Fatma Romeh M., and Shiferaw Gurmu. "The impact of female education on fertility: a natural experiment from Egypt." *Review of Economics of the Household* (2016): 1-32.

Assaad, Ragui, and Mohamed Saleh (2016). "Does Improved Local Supply of Schooling Enhance Intergenerational Mobility in Education? Evidence from Jordan." *The World Bank Economic Review* (2016).

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2008). "Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births." *The Economic Journal* 118.530: 1025-1054.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocıo Titiunik. "Regression Discontinuity Designs Using Covariates." (2018). Working Paper.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. (2016) rdrobust: Software for regression-discontinuity designs. Stata Journal 17: 372-404.

Campante, Filipe R., and Davin Chor (2012). "Why was the Arab world poised for revolution? Schooling, economic opportunities, and the Arab Spring." *Journal of Economic Perspectives* 26(2), pp. 167-88.

Fan, Jianqing, and Irène Gijbels (1996). "Local Polynomial Modelling and Its Applications." *Chapman & Hall, London*.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica* 69.1 (2001): 201-209.

Hodges, Rebecca McLain. *Transformation of Sympathies: Gendered Mediation of Jordanian Education Reform for a Knowledge Economy*. Washington University in St. Louis, 2015. ProQuest Number: 3725247

Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142.2 (2008): 615-635.

Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.0 [dataset]. Minneapolis, *MN: IPUMS, 2018*.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos (2004). "Does education improve citizenship? Evidence from the United States and the United Kingdom." *Journal of Public Economics* 88(9): 1667-1695.

OAMDI, 2017. Jordan - Employment and Unemployment Survey, EUS. Years: 2006, 2007, 2008, 2009, 2010, 2011, 2013, 2014, and 2016. Harmonized Labor Force Surveys (HLFS), Version 1.0 of Licensed Data Files; EUS - Department of Statistics (DOS), The Hashemite Kingdom of Jordan. Egypt: Economic Research Forum (ERF).

Oreopoulos, Philip (2006). "Estimating average and local average treatment effects of education when compulsory schooling laws really matter." *The American Economic Review* 96(1): 152-175.

Oreopoulos, Philip (2007). "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling." *Journal of Public Economics* 91(11): 2213-2229.

Oreopoulos, Philip, Marianne E. Page, and Ann Huff Stevens (2006). "The intergenerational effects of compulsory schooling." *Journal of Labor Economics* 24(4): 729-760.

Pischke, Jörn-Steffen, and Till Von Wachter (2008). "Zero returns to compulsory schooling in Germany: Evidence and interpretation." *The Review of Economics and Statistics* 90(3): 592-598.

Powdthavee, Nattavudh (2010). "Does education reduce the risk of hypertension? Estimating the biomarker effect of compulsory schooling in England." *Journal of Human Capital* 4(2): 173-202.

Robbins, Philip (2004). A History of Jordan. Cambridge: Cambridge University Press, 2004.

Silles, Mary A (2009). "The causal effect of education on health: Evidence from the United Kingdom." *Economics of Education Review* 28(1): 122-128.

Silles, Mary A (2011). "The effect of schooling on teenage childbearing: evidence using changes in compulsory education laws." *Journal of Population Economics* 24(2): 761-777.

Thistlethwaite, Donald L., and Donald T. Campbell. "Regression-discontinuity analysis: An alternative to the ex post facto experiment." *Journal of Educational psychology* 51.6 (1960): 309.

Van Kippersluis, Hans, Owen O'Donnell, and Eddy Van Doorslaer (2011). "Long-run returns to education does schooling lead to an extended old age?." *Journal of Human Resources* 46(4): 695-721.

World Bank (2018). World Development Indicators. https://data.worldbank.org/indicator/ Accessed: June 2018

Chapter 3. Sweet Child O' Mine: Son Preference in the United States

3.1 Introduction

Son preference has been found in many Asian countries like China, India, Korea, and Vietnam. (Arnold and Liu, 1986; Gupta, 1987; Aly et al., 1991; Haughton and Haughton, 1998; Edlund, 1999; Das Gupta et al., 2003; Jayachandran and Pande, 2017). Studies show that parents in these countries exhibit favoritism for sons over daughters both before and after birth, with consequences ranging from higher fertility, to imbalanced sex ratios, to gender discrimination. Some research has begun to extend the analysis to other countries, such as South Africa, Canada, UK, and the U.S. (Gangadharan and Maitra, 2003; Dubuc and Coleman, 2007; Dahl and Moretti, 2008; Abrevaya, 2009; Almond, 2013).

In the case of the U.S., existing studies have yielded some inconsistencies.[58] Dahl and Moretti (2008) find that parents in the U.S. prefer boys, while Abrevaya (2009) finds no evidence of son preference in the United States in aggregate. Instead, son preference is only documented for Chinese and Indian immigrant's sub-populations. More recently, Blau et al. (2017) show that having a daughter first is associated with lower fertility for the overall population. Much of the difference may stem from methodology. For instance, Dahl and Moretti (2008) present indirect evidence of son preference among parents in the U.S. by estimating the effects of child gender on mothers' marital status and family structure using the 1960 to 2000 U.S. Censuses. They find that a first-born girl is less likely to live with her father relative to a first-born boy due to the mother's lower probability of

[58] In Canada, the first-generation immigrants from South and East Asia tend to have stronger son preference than the second-generation immigrants, but when the first-generation immigrants are separated into two groups, arriving before and after age 18, there is no significant difference among them for son preference (Almond, 2013).

getting married, higher probability of being divorced, and a father's lower likelihood of obtaining custody of daughters. Abrevaya (2009) provides evidence of gender selection among parents who migrate from Asia like China, India, and South Korea using vital statistics from Federal and California from 1970 to 2005 and census data (1980, 1990, and 2000) to estimate the boy-birth percentage at later births according to the dimensions of parents' source countries and previous genders.

Even though son preference is inconsistent at the aggregate level, evidence that immigrants from Asian countries exhibit son preference in the U.S. is more robust. Despite the relative size of this group, accounting for 7% of the overall population in this study, little research attention has been devoted to examining the formation and the extent to which these preferences persist within U.S. society.

On the one hand, studies show that son preferences can change with the social values over time in Japan and South Korea. (Kureishi and Wakabayashi, 2011; Choi and Hwang, 2015) Similarly, studies like Blau et al. (2008) find intergenerational assimilation of immigrants to native levels for fertility, education, and labor supply. Bleakley and Chin (2010) find that greater English proficiency of immigrants, estimated by using immigrants' age at the time of arrival in the U.S. as IV, leads to more assimilation to U.S. behaviors, including a higher probability of divorce and intermarriage and a less likelihood to live in ethnic enclaves. Thus, we might expect that the son preference among Asian immigrants may change with the process of their assimilation in the U.S.

On the other hand, evidence suggests that not all culture characteristics of immigrants can be so easily assimilated even with longer duration of residence in the U.S. For instance, Hicks et al. (2015) find that gender norms of behavior in the allocation of household is acquired in childhood, and once established, is relatively unaffected by

duration of exposure to U.S. culture. Therefore, there is a critical period for the age of migration which affects the effect of assimilation on the persistence of gender norms of immigrants, and also might affect immigrants' other cultural markers with applications.

I present two analyses in this article using American Community Survey data from 2012 to 2016. First, I estimate whether parents in the U.S. have son preference in aggregate and across immigrant populations, and then I extend the analysis to higher birth orders for large size families, and further I extend the estimation to the indirect mechanism through the effects of the gender of children on mothers' marital status. Second, I analyze whether son preference among immigrants from South and East Asian countries are assimilated by the local community in the U.S. over time, contrasting groups of first-generation and higher order generation immigrants, as well as contrasting son preference across mothers with different age of arrival to the U.S and with different duration of residence in the U.S. before the birth of their first child.

In the analysis, I use whether the family had another child within six years as a dependent variable to test for son preference via a fertility stopping mechanism. In theory, if parents prefer sons rather than daughters, then the likelihood of having additional children should be higher for families with daughters only than for families with sons only. I find that for both small size and large size families, there is no evidence of son preference in the U.S. at large. Additionally, I use whether the next child is a boy as another dependent variable to test son preference through a gender selection mechanism.[59] Results again show that there is no significant evidence of gender selection in aggregate in the U.S.

[59] The means for gender selection include three types of options, the first type covers infanticide and child abandonment which usually occur after children's birth and the gender of children is observed; the second type is abortion which happens before the birth of a baby but after the formation of a fetus. Abortion access is available mainly due to the introduction of ultrasound and amniocentesis, which can identify the gender of a fetus, in the 1970s; the last type includes in vitro fertilization (IVF) and sperm sorting which are

As robustness, I run a proportional hazard function model as it provides additional information on son preference by exploiting the complete set of variation in birth spacing across children. Another advantage of this model is it doesn't require a complete family sample, so mothers in the sample are not restricted to the age between 18 and 40 (Haughton and Haughton, 1998). The results of Cox Proportional Hazards model are consistent with the previous analysis.

Next, I turn to son preference among immigrants who migrated from Asian countries. I start from the estimation of son preference among the first-generation immigrants and then separate sub-populations based on country of origin. I also construct a control group which contains immigrants from G7 countries.[60] I find that, relative to natives and relative to G7 immigrants, parents who migrate from Asian countries show stronger son preference, both among small and large size families, and this occurs through both mechanisms previously described – fertility stopping and gender selection. Furthermore, the results show that parents from China, India, and Vietnam have son preference while those from Japan and Korea do not. When I consider the higher order generation immigrants, only parents from India still have son preference, while immigrants from China and Vietnam are completely assimilated by the external environment.

Finally, using a differences-in-differences methodology, I compare the fertility as well as gender selection decisions of immigrants with daughters only and sons only arriving in the U.S. before and after the critical period of assimilation (here taken to be around age 9).[61] I find that fertility and gender selection decisions are particularly skewed along parents

performed before pregnancy. They are processed on the fertilized embryos to obtain the desired gender and then transferred into mother's uterus to form the fetus. (Abrevaya, 2009; Almond, 2013)

[60] G7 countries consist of Canada, France, Germany, Italy, Japan, the United Kingdom and the United States.

[61] The "critical period" of assimilation is derived from the linguistic hypothesis, the critical period of language acquisition. It suggests that people are more easily to be proficient in a language and its grammatical structure

who migrated after the critical period. This strongly suggests that exposure to home country norms regarding son preference in a period of development at a young age is critical for developing son bias.

The reminder of the paper is organized as follows. Section 2 discusses the sample and data summary. Section 3 explains the regression model. Section 4 discusses the results. Section 5 extends the regressions to the use of hazard model for the completed sample, indirect evidence for son preference in the U.S., and the formation and assimilation of son preference. Conclusions are given in Section 6.

3.2 Sample and Data Summary

The American Community Survey (ACS) is conducted yearly and includes information on education, housing, jobs, immigration and other demographic characteristics.[62] Data are available at the individual level with the identifiers linking to the structure of family and household. I employ the 2012-2016 ACS 5-year Public Use Microdata Sample (PUMS) data. It contains all cases from the previous 1% ACS samples for 2012, 2013, 2014, 2015, and 2016, and reflects approximately 5% of the US population.[63]

I define a family in this analysis as a married mother with an eldest child at the age of 12 or younger. To avoid the problem that some of children may have left the household, I limit mother's age to between 18 and 40 in the sample.[64] Both of these choices

in childhood, so people who migrate early in life to the U.S. tend to learn and speak English more fluently relative to those immigrants arriving later in life. This hypothesis has been applied in many studies to analyze the process of assimilation of immigrants. (Bleakley and Chin, 2004, 2010; Hicks et al., 2015).

[62] The American Community Survey includes questions not asked in the regular census. Response to the survey is required by law Title 18 U.S.C Section 3571 and Section 3559, which amends Title 13 U.S.C. Section 221.

[63] The primary difference between the 2010 census and the 2012-2016 ACS file for the purpose of this paper is the availability of detailed information on immigration and household structure.

[64] Results are still robust when I use higher tail of 45 or 50.

dramatically limit the chance that sample selection and attrition would drive the results observed.

The resulting sample size for families with at least one child is 490,858.[65] Among that, household size, on average, is 3.99, and the sample size for households with only one child is 175,675, with two children are 213,854, and with more than two children is 101,329. Summary statistics are presented in Table 11. 62% of mothers are in the labor force and are 92% fathers working. Almost 93% of mothers attain a high school education or above and 46% of mothers have at least a bachelor's degree. Education attainment for fathers, at 89% and 39%, are lower than mothers. In the sample, 30% of families have only daughters, and 32% of households have only sons. The rest, 38% of families, have both daughters and sons. 20% of families have a foreign-born mother, and 35% of them are from the South, East, and Southeast Asia.

Individual age is reported for each household member. Using this information, I construct birth order for children in a family and record the gender of the child at each order of birth. Further, I additionally construct variables for whether the family had another child and whether the family had another male child.[66]

The ACS provides information on each individual's place of birth, allowing us to define the first-generation immigrants and where they are from. First, I decompose the sample into natives and first-generation immigrants to estimate son preference separately for these two groups. Second, I break the first-generation immigrant groups into people

[65] The unit of observation is household.

[66] In this article, I concern both the two children and three children families. When I estimate the son preference for the second child, families with twins of the first two children and triples are excluded out. When I estimated the son preference for the third child, families with twins of the second and third child as well as triples are excluded out.

who migrated from G7 countries and South, East, and Southeast Asian countries.[67] As the ACS also contains information on each person's ancestry and race, I define higher-order generations of immigrants of South, East, and Southeast Asian (hereafter SE Asian) as natives who claim their ancestry are from South, East, and Southeast Asia or as individuals who speak a specific Asian language at home or who claim themselves having at least one third of race of a specific Asia country.

Finally, for first-generation immigrants, year of entry is available. I use this variable to test whether the son preferences among SE Asian immigrants change with age at the time of arriving in the U.S., which measures the extent of social assimilation of immigrants.

3.3 Analysis

3.3.1 Baseline Regression

I begin the analysis within the following econometric framework:

$$Y_i = \beta_0 + \beta_1 AllGirls_i + X_i\Gamma + \varepsilon_i \quad (1)$$

where i denotes married women between the age of 18 and 40 with at least one child, $AllGirls_i$ is a dummy variable indicating the gender of previous children as being all female, X_i is a vector of control variables, ε_i is the error term, βs are the coefficients, and specifically, β_1 is the parameter of interest.

I consider two dependent variables of interest: $Fertility_i$, indicating whether family i had additional children within six years, and Male child_i, indicating whether the next child is a son.[68] For the "had second child" sample, I include families with at least one

[67] All countries and regions in South, East, and Southeast of Asia are considered once they are recorded in ACS. It includes China, Hong Kong, Taiwan, Japan, Korea, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam, Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, and Sri Lanka.

[68] By using five-year or seven-year window, I get similar results for the regressions.

child, so if they had a second child within six years after the first child's birth, $Fertility_i$ equals 1, otherwise it is 0. Similarly, in the "had third child" sample, I consider families with two and above children. If they have a third child within six years of the birth of the second child, then the dependent variable, $Fertility_i$, would take the value 1. For another dependent variable, Male child_i, it takes the value 1 when the additional child is male.[69] The predicted values of having another child and whether the next child is a boy estimated by the Linear Probability Model (LPM), are predominantly ranged between 0 and 1.[70]

The parameter of interest is the coefficient on $AllGirls_i$, an indicator variable assigned 1 when the previous children are all girls in family i. The coefficient indicates the difference in Y_i between families with all girls and those with at least one son. Arguably, it should provide an upper bound of the effects of son preference on family size because some families may have additional children to balance the gender pool of their children – an effect not due to son preference. To rule out that concern, in some specifically, I additionally control for Mix_i, which indicates the gender of previous children contains both boy and girl, in the following regression equation.

$$Y_i = \beta_0 + \beta_1 AllGirls_i + \beta_2 Mix_i + X_i\Gamma + \mu_i \quad (2)$$

Once this additional indicator is included, the omitted category is that for a group of families with only boys. Thus, the coefficients on $AllGirls_i$ and Mix_i indicate the difference between families with all girls or families with children of both sexes in comparison to the fertility of those with only sons.

The vector of control variables, X_i, are standard to the literature on fertility and son preference in general, and contain variables that relate to the mother, household

[69] There is no six-year window for dependent variable, Male child_i.
[70] 98% of the predicted values are between 0 and 1.

variables, and geographic variable.[71] The maternal variables include whether the mother has high school and above education, whether the mother has at least a bachelor's degree, mother's age, and mother's age at marriage. Household variables include family income, father's education attainment captured by dummy variables for high school and above as well as bachelor's degree and above, and whether the head of household is male. I also include the geographic indicator, state dummies, to control for an influence of regional factors on families' son preference, like differences in state level access to abortion or differences in regional locations of ethnic enclaves.

I use a linear probability model with robust standard errors to estimate the above equations since the two dependent variables are both dummy variables. The coefficients of control variables can be interpreted as the change in the probability of having another child within the six-year window and the next child being a boy respectively. Meanwhile, I run a Cox Hazard regression as a robustness check and also provide additional evidence for son preference. For the hazard function, I don't limit the age of the mother in the sample to between 18 and 40. Thus, I have a larger sample for the hazard regression.

3.3.2 Immigrant Assimilation

Studies on son preference among Asian immigrants typically produce consistent and positive results, indicating son preference in the US for these groups. A set of important unanswered question is when during life one develops this son preference, whether it can be still developed in the U.S. among local enclave populations, and whether son preference among Asian immigrants vanishes when they have resided in the U.S. for a long time.

[71] Haughton and Haughton (1998) illustrate that those factors also affect families' fertility and need to be controlled for in the regressions.

For the estimation of immigrants' assimilation of son preference, I propose two sets of analysis. The first one examines intergenerational assimilation between generations of immigrants. I again compare the change in the probability of having another child and the next child being a boy if the previous children are all girls relative to all boys – now taking the difference across first-generation parents who migrated from Asian countries against those of second and higher generations in the U.S.

The second channel considers the development of son preference and assimilation at the individual level, exploiting variation in one's age at arrival to the U.S. In order to do this, I estimate a differences-in-differences regression according to the following equation to capture the difference in son preference across immigrants arriving before and after this critical age period.

$$Y_i = \beta_0 + \beta_1 AllGirls_i + \beta_2 Mix_i + \beta_3 Age\_AtArrival_i + \beta_4 AllGirls_i * Age\_AtArrival_i + \beta_5 Mix_i * Age\_AtArrival_i + X_i\Gamma + \eta_i \quad (3)$$

$Age\_AtArrival_i$ is a dummy variable, indicating whether or not mother i migrated to the United States after the cutoff age. I follow previous immigrant assimilation literature and choose nine years old as the cutoff age for this analysis.[72] The coefficient of interest, that in front of the interaction term of $AllGirls_i$ and$Age_AtArrival_i$ can be interpreted as the change in the probability of having another child (or of the next child being male) should the previous children of family i be all girls specifically for mothers who have arrived at the age of nine and above in comparison to that change in probability for those who arrived at a younger age.

[72] Bleakley and Chin (2004, 2010) and Hicks et al. (2015) use nine years old as the "critical period" cutoff. I get similar results when I choose eight or ten years old as the cutoff age.

3.4 Results

3.4.1 Results for Basic Regression

Table 12 reports the influence of son preference on fertility for the full sample of families with oldest child 12 years of age or younger in the U.S., as well as for the sample of families where the mother is either native-born, or a first-generation immigrant, a G7 immigrant, or a South, East, and Southeast Asian immigrant.[73] Panel A of Table 12 provides the results of preference for at least one son, outcomes of regression equation 1, for the above five samples. For families with at least one child, the coefficient of interest in front of all girls for parents in the U.S. is -0.002 and insignificant, indicating no strong influence of initial child gender on the decision to have a second child.

The coefficient for the sample of natives is actually negative and significant. This means that the probability of having a second child within the next six years is 0.4 percentage points lower if the first child of the family is a daughter relative to a son. For first-generation immigrants, this pattern is reversed. Having only a girl leads to a positive and significant increase in the odds of having a second child. Specifically, for the sample of immigrants who migrate from South, East, and Southeast Asia, the coefficient is 0.023 and significant at 1%. This implies that immigrants from Asian countries are 2.3 percentage points more likely to have another child if they have a previous girl to a boy. Comparatively, the coefficient on all girls for immigrants from G7 countries is negative and not significant. Thus, I conclude that families who migrate from SE Asian countries are more likely to have a second child if the previous child is a girl rather than a boy.

[73] The complete regression results with coefficients on all covariates are available upon request.

For families considering a third child, the coefficients on all girls are all positive and significant at 1% across all samples. It seems like parents in the United States have the incentive to have a third child if the previous children are all girls, but people would argue that families with all girls tend to have a higher probability of having another child relative to families with at least one son simply due to preference of gender balancing but not son preference. To exclude out the effect of this concern, I run the regression of equation 2 additionally controlling for the dummy variable mix, indicating families with both daughter and son.[74] The coefficients on all girls are again negative for the full sample and natives, and it is significant at 1% for natives.

For G7 immigrants, the coefficient becomes positive but not significant. For first-generation immigrants and SE Asian immigrants group, the coefficients are still positive and significant at 1%. Furthermore, the magnitudes are even larger than the case of families with at least one child. These results show that, son preferences among most households in the U.S. are mainly driven by a preference of gender balancing. On the other hand, immigrant families with previous children being all girls are more likely to have a third child than families with two boys first, a phenomenon that is especially true for SE Asian immigrants.

Table 13 follows the same empirical strategy, this time for another dependent variable, whether the next child is male.[75] For families with at least one child, the coefficients in front of all girls are all negative for the samples of families in the United States, natives, first-generation immigrants, and G7 immigrants, and they are significant at 1% for the first two groups. For the sample of SE Asian immigrants, the coefficient is

[74] The coefficients in front of mix are all negative and significant for all samples, providing evidence for gender balancing.
[75] The complete regression results with coefficients on covariates are available upon request.

positive but not significant. This implies that the first-generation of Asian immigrants did not exhibit gender selection for the second child.

For families with at least two children, the coefficients on all girls are all positive and significant at 1% for those samples. However, this appears to be mostly driven by gender balancing preferences for native born households. Coefficients in front of all girls are positive and significant at 10% to 1% for immigrants. G7 as well as SE Asian immigrants are particularly more likely to have a son as their third child if the previous two children are all girls relative to all boys.

I conclude that there is little evidence for son preference among natives through either fertility stopping decisions or from gender selection in the U.S. At the same time, families who migrate from South, East, and Southeast Asia countries are more likely to have additional children if the previous children are all girls than all boys, and appear to influence the gender of the third child. A broader pool of G7 immigrants does not exhibit son preference to nearly the same degree.

3.4.2 Regional Disaggregation

To better understand the son preference among SE Asian immigrants, I decompose the sample of SE Asian immigrants into families who migrate from a specific country in South, East, and Southeast Asia. With the concern of previous literature and sample size, I run regressions on the sample of immigrants who are from China, Japan, Korea, Philippines, Vietnam, and India respectively as shown in Table 14 and 15. For families who migrate from China, Vietnam, and India with at least one child, they are 4.3, 3.4, and 2.9 percentage points more likely to have a second child if the family had a girl first relative to have a boy first. Controlling for gender balancing, when the previous two children are all girls, families from China and Vietnam both exhibit a higher propensity to

have a third child than families with all boys. Indian immigrants still hold the similar change in the probability of having another child for families with girls only to boys only when families with at least two children as families with at least one child.

Interestingly, first-generation immigrants who are from Japan, Korea, and Philippines do not exhibit a significant preference to have an additional child if the previous children are all daughters rather than all sons. Panel A of Table 15 shows that there is no significant evidence for gender selection among first-generation immigrants who are from those six countries to have a son as their second child. While, immigrants who are from China, Vietnam, and India have higher, 6.1, 4.2, and 2.1 percentage points, the probability to have a male child if the previous two children are all girls rather than all boys.

To most easily compare the magnitude and significance level of son preference for different samples, I draw the coefficients on all girls with confidence interval for regressions as shown in Figure 16. Panel A and B provide the coefficients on all girls both for families with at least one child and those with at least two children. It can be clearly seen from the regression results that immigrants from China, Vietnam, and India have stronger son preference than natives and immigrants from other countries. From this section, I conclude that there is no strong evidence for son preference among immigrants who are from Japan, Korea, and Philippines, while immigrants who are from China, Vietnam, and India exhibit son preference and may also practice some gender selection for higher birth order children when faced with no male births.

3.4.3 Results for Immigrant Assimilation

Tables 16 and 17 provide results for the fertility stopping and gender selection hypotheses for higher order generation immigrants who are from the South, East, and

Southeast Asian countries. In panel A of Table 16, it is immediately apparent that there is no significant change in the probability of having a second child between families with a first girl and families with a first boy. This coefficient becomes positive for larger size families. When I control the preference for gender balancing, higher order generation immigrants who are from Asian countries, in aggregate, have 1.8 higher percentage points to have a third child if they have two girls first rather than two boys, but the effect is not significant for the pooled sample.

When I disaggregate the analysis again to country level, families whose ancestors are from India are 10.9 more percentage points to have a third child if the previous children are all girls than all boys, and the coefficient on all girls is significant at 1%. This effect massively dwarfs those from previous regressions.

For the results of gender selection behavior as shown in Table 17, higher order generation immigrants again do not significantly influence the gender of their second child even if their first child is a daughter. However, families whose ancestors are from India still remain an outlier, with a strong preference to have a son should the first two births be girls. The coefficient on all girls is 0.076 and significant at 5%.

Comparing the results of higher order generation immigrants to the results of first-generation immigrants, I find that immigrants who are from Asian counties tend to have a lower change of probability in having additional children if they have all girls with higher order generation than with first-generation. Assimilation is most prevalent among immigrants who are from China. First-generation immigrants have the strongest son preference among other Asian immigrants, however, I don't find any significant evidence of son preference for the higher order generation immigrants. Meanwhile, I don't find

significant evidence of son preference among both first-generation and higher order generation immigrants who are from Korea and Philippines.

Indian immigrants are again a clear outlier.[76] The coefficients on all girls are larger for higher order generation immigrants than the first-generation immigrants. This discrepancy is surprising and suggests an interesting avenue for future investigation. It is mainly from the unusual strong son preference among higher order generation Indian immigrants. The plausible explanation is provided at the end of this section.

To further understand the process of formation and assimilation among immigrants, I run a differences-in-differences regression on the sample of first-generation immigrants who are from SE Asian countries, by adding the dummy variable Arrived at age 9+, whether immigrants have arrived at age nine and above, and the interaction term of it with all girls as described in equation 3.

Tables 18 and 19 provide results for fertility stopping decision and gender selection behavior respectively. The coefficients on arrived at age 9+ are significantly negative. This implies that families who migrate at an older age tend to have fewer children than families who arrived at an earlier age. The sum of the coefficients in front of arrived at age 9+ and the interaction term of it with all girls are still negative for all samples.[77] This result is consistent with the difference between the first-generation and higher order generation Asian immigrants. Simply comparing the group of arriving at younger than nine years old

[76] I run a differences-in-differences regression on the two samples as following: $Y_i = \beta_0 + \beta_1 AllGirls_i + + \beta_2 Mix_i + \beta_3 HigherOrderG_i + \beta_4 AllGirls_i \times HigherOrderG_i + \beta_5 Mix_i \times HigherOrderG_i + \beta_6 X_i + \theta_i$, for both dependent variables. The coefficients in front of the interaction term of all girls and higher order generation are 0.104 and 0.062 respectively, and they are significant at 1% and 5%. However, when I run the above regressions for sample of immigrants who are from China, the coefficients in front of the interaction term of all girls and higher order generation for the dependent variables, fertility and male child, are 0.16 and -0.033, and none of them is significant at 10%.

[77] The combination of coefficients in front of Arrived at age 9+ and the interaction term of it with all girls is the coefficient for families who migrated at age nine and above with all girls.

to above, the group of arriving at age nine and above has a smaller average number of children than below.[78] The explanation for this could be the pressure from surviving in the United States or selection into the age of migration on the basis of economic outcomes not captured perfectly by our set of controls.[79]

Families who migrate at an older age may experience more pressure of communication, relationship, and different culture than those who arrive at a younger age. Column 1 of Tables 18 and 19 provide the results of the differences-in-differences regression on the sample of immigrants who are from the South, East, and Southeast Asia in aggregate. The coefficients on the interaction term of all girls and arrived at age 9+ are both positive and significant for the dependent variable fertility decision.[80] For another dependent variable, gender selection, the coefficient on the interaction term is 0.055, and it is significant at 1% in families with at least one child. For families with at least two children, the coefficient is positive but not significant.

As such, immigrants who are from Asian countries, in aggregate, have stronger son preference if they arrived at age nine and above than they arrived at younger. This finding could also be consistent with exposure to home country norms influencing later life fertility decisions, assuming that early years are formative years regarding these sets of preferences.

Figure 17 and 18 flexibly plot regression adjusted coefficient estimates on all girls across mothers' age of migration within first-generation immigrants from Asian countries with at least one and two children. The results are consistent with the choice of the critical

[78] For the samples of first-generation and higher order generation immigrants who are from South, East, and Southeast Asia countries, I also find a larger mean of number of children for the higher order generation.
[79] For example, maybe initially they spend more time and money on surviving in the new countries. They get married later, spend less time on children, and have stronger incentive to integrate into local community.
[80] They are significant at 1% and 5% respectively.

period cutoff as the coefficients become positive and significant only among families who migrate at later in their life. These flexible plots suggest between 7 and 11 is an appropriate age to be used as the cutoff.

There is again heterogeneity in the results for the sample at the country level as shown in column 2 to 5 of Tables 18 and

9. The change in the probability of having a third child for immigrants who are from China is 17.1 more percentage points in families whose previous children are all girls and who migrated at the age of nine and above relative to those with all boys and arrived at the age of nine and above. Moreover, Chinese immigrants who had girls only and arrived at the age nine and above have a higher probability of having another male child than families had boys only and migrated at the age of nine and above. Immigrants who are from India and Vietnam also have the stronger preference to have a third child if the families who migrated at the age of nine and above with girls only relative to those arrived at the age of nine and older with boys only. However, they don't have significant bias on the gender of the second and third child even the previous children are all daughters. There is no significant evidence of son preference among immigrants who are from Korea and Philippines for the different groups of the gender of previous children and arriving age.

Combining the results with the variation in different generations and arriving age, I find evidence for the assimilation of son preference among SE Asian immigrants. Families who are from SE Asian countries tend to have stronger son preference among the first generation. They also exhibit strong son preference when they arrived at the age of nine or above. The assimilation among Chinese immigrants stands out in particular. They have the strongest son preference for the first-generation immigrants among the analyzed countries, while, I don't find any significant evidence of son preference among

higher-order generation immigrants. The assimilation among Vietnamese immigrants is also obvious. However, the trend is different for Indian immigrants. They have stronger son preference among higher-order generation immigrants than the first-generation immigrants. One plausible explanation may be the higher proportion of marriages among the same ethnic group for Indian immigrants than immigrants from other ethnic groups. The high proportion of marriages among Indian immigrants facilitate those families keep their own tradition and social norm which includes son preference at home. In Table 20, 92% of first-generation Indian immigrant mothers get married to first-generation Indian immigrant fathers.[81] This statistic is consistent with the plausible explanation.

3.5 Extensions

3.5.1 Cox Hazard

As a robustness check, I consider a more detailed analysis of birth spacing. Cox hazard regression results, presented in Table 21, are broadly consistent with the results of LPM regressions. The coefficients on all girls for the full sample and natives are negative and significant. The corresponding hazard ratios are less than 1. These ratios indicate that families whose first child is a girl are 0.7% and 0.9% less likely to have another child than those whose first child is a boy.

For families who migrated from Asian countries, parents whose first child is a daughter are 0.9% more likely to have another child than those whose previous child is a son, but the coefficient is not significant. When I consider higher parity of children, even though the coefficients are no longer precisely estimated, they still produce the same sign as previous regressions.

[81] For first-generation immigrant Chinese and Vietnamese, the proportions are 72% and 75%.

3.5.2 Effects of the Gender of Children on Mother's Marital Status

Even though we don't find evidence of strong forms of son preference among natives in the U.S. through fertility stopping decision and gender selection, it does not mean there could not be weaker evidence of son preference. According to Dahl and Moretti (2008), son preference in the U.S. can be seen in parents' marital status, with marriage more commonly occurring before and after male births.

In Table 22, I run regressions in which the dependent variable is a dummy equal to 1 if a mother with at least one child is divorced, separated, or never married, and 0 if the mother is married at the time of the survey. This is regressed on the gender of previous children with the complete set of covariates from equation 3 as well as a control for number of children and for families with one child or more. The indicator of the gender of previous children labeled 'all girls' is set equal to 1 if the previous children are all daughters.

The estimated coefficients on all girls indicate the change in the probability of divorce if families have girls only relative to they have at least one son. The coefficients are all positive, and for the full sample, natives, as well as first-generation immigrants, they are significant at 1% to 5%. This indicates that U.S. households, consistent with the previous analysis, are either more likely to marry, or less likely to separate after having sons.[82] Even though the magnitudes, 0.5, 0.6, and 0.2 percentage points, seem like small, percentage effects are actually very large, at 9.5%, 11%, and 4.4%, and are thus quite non-negligible.[83] Thus, I can conclude that son preference in the U.S. is not obvious, and it is

[82] I cannot precisely define the time of divorce. The estimated results maybe the outcomes of divorce but not the reason, for example, fathers are more likely to get the custody of sons but not daughters when they get divorced. In this case, it still shows at least fathers' son preference in the U.S.

[83] There are 5.28% mothers with at least one child are divorced, separated, or never married in the U.S. There are 5.47% native mothers with at least one child are divorced, separated, or never married. There are

perhaps more likely to operate via fathers rather than mothers among the native born population. Or I can conclude that among normal marriage, there is no significant evidence of son preference among parents in the U.S., but out of normal marriage, fathers prefer sons rather than daughters. Although the estimates for the subsamples of G7 immigrants and SE Asian immigrants are of the same sign, they are imprecisely estimated.[84]

3.5.3 Formation and Assimilation of Son Preference

The results of Tables 18 and 19 as well as Figures 17 and 18 suggest that individuals develop their son preference early in life, most likely about at the age of 7 to 11 years.[85] A natural next question is whether son preference, once developed, is persistent. Figures 19 and 20 plot the coefficient on all girls as a function of the duration of time spent in the U.S. before the birth of the first child.[86] The logic here is to examine whether the extent of son preference among first-generation Asian immigrants may change with the duration of time resided in the U.S.

In these specifications, I additionally control for age of migration since we have already shown that it matters for the formation of son preference. As can be seen from the figures, longer duration of residence in the U.S. has little bearing on the degree of son preference exhibited in these households. In Figures 19 and 20, the coefficients on all girls are consistent across the duration of time spent in the U.S.[87] This suggests that first-

4.57% mothers who are first-generation immigrants with at least one child are divorced, separated, or never married.

[84] The proportions for G7 and SE Asian immigrants are 2.22% and 3.77% respectively.

[85] This period is consistent with the period of formation of other language related norms. (Bleakley and Chin, 2004; Hicks et al., 2015)

[86] Duration of time spent in the U.S. equals to age at time of immigration minuses mother's age at the birth of her first child.

[87] When I choose 18 years old as the age cut off for regressions in Table 8 and 9, the coefficients on all girls are smaller than the previous results and insignificant. This result is consistent with Almond et al. (2013)

generation Asian immigrants in aggregate are unlikely to have these particular cultural preferences altered by exposure to U.S. culture.

3.6 Conclusion

This paper studies several populations of families in the U.S., including both natives and immigrants, to explore the repercussions of son preference. In order to do this, I analyze the effects of the gender of previous children on families' further fertility decision and gender selection behavior. As a novel addition, this article explores the formation and malleability of these preferences – studying assimilation of son preference among Asian immigrants through the variation of the age of migration and intergenerational transmissions.

I find that there is no robust evidence for son preference among natives in the U.S., but there is indirect evidence from the effects of the gender of children on mother's marital status to support the existence of son preference among natives. I also show that the change in the probability of having additional children and the next child being male is significant and positive for Asian immigrants if their previous children are all girls relative to all boys. First-generation immigrants again exhibit stronger son preference than higher order generation immigrants who are from Asian countries.

Moreover, Asian immigrants who arrived at the age of nine and above are significantly more likely to have another child if the gender composition of previous children is female than families who migrated at the age of nine and above with previous children being all male. Son preference among SE Asian immigrants is developed in early of life, and once established, it is kind of persistent with the first-generation Asian immigrants who migrated after the age of nine. However, I find significant evidence of assimilation of son preference among intergeneration Asian immigrants. The most

obvious assimilation occurs for Chinese immigrants. They exhibit the strongest son preference among analyzed Asian countries with the first-generation immigrants, yet this disappears for higher-order generations. Indian immigrants show the opposite trend. Higher order generation of Indian immigrants has stronger son preference than the first generation. These results suggest interesting avenues for future research.

Table 11. Summary Statistics

Household Level	mean	sd
Panel A: Demographics / Education Statistics		
Household size	3.99	0.98
Family income	98,810	89,883
Mother in labor force	0.62	0.49
Mother's age at marriage	27.87	5.86
Mother's education high school and above	0.93	0.25
Mother's education bachelor's degree and above	0.46	0.50
Father in labor force	0.92	0.27
Father's education high school and above	0.89	0.31
Father's education bachelor's degree and above	0.39	0.49
Panel B: Country of Birth / Ancestry		
First-generation immigrants	0.20	0.40
Immigrants from East, SE Asian countries and regions	0.07	0.26
Immigrants from China	0.01	0.10
Immigrants from Japan	0.003	0.05
Immigrants from Korea	0.01	0.07
Immigrants from Philippines	0.01	0.10
Immigrants from Vietnam	0.01	0.08
Immigrants from India	0.02	0.15
Higher order generation immigrants (SE Asian countries)	0.02	0.14
Panel C: Fertility		
Mother's age of first birth	26.61	4.74
Number of children	1.91	0.89
% of household with all girls	0.30	0.46
% of household with all boys	0.32	0.47
% of household with mix	0.38	0.49

Notes: The sample covers families with married mothers between age 18 and 40, and with at least one child and the oldest one is 12 years old or younger in the United States. The unit of observations is the household, and there are 490,858 households in the sample.

Source: American Community Survey from 2012 to 2016.

Table 12. The influence of son preference on fertility in the U.S.

	Fertility				
	Full Sample	Natives	First-generation Immigrants	G7 Immigrants	SE Asian Immigrants
Panel A: Preference for at least one son					
Had Second Child [(a)]					
All girls	-0.002	-0.004***	0.007**	-0.005	0.023***
	(0.001)	(0.002)	(0.003)	(0.012)	(0.005)
Had Third Child [(b)]					
All girls	0.032***	0.030***	0.041***	0.043**	0.050***
	(0.002)	(0.002)	(0.004)	(0.017)	(0.006)
Panel B: Son preference controlling for gender balancing effects					
Had Third Child					
All girls	-0.004	-0.008***	0.015***	0.018	0.031***
	(0.002)	(0.003)	(0.005)	(0.020)	(0.008)
Mix	-0.056***	-0.060***	-0.039***	-0.038**	-0.029***
	(0.002)	(0.002)	(0.004)	(0.016)	(0.006)

Notes: The dependent variable, fertility, indicates having another child within six years of previous child's birth. (a). "Had second child" represents the sample of families with at least one child and may have additional children later. The unit of observations is the household, and the sample size for each group is 480,360; 383,683; 96,677; 6,443; 34,592 respectively. (b) "Had third child' represents the sample of families with at least two children and may have additional children later. The sample size for each group is 308,378; 248,195; 60,183; 3,970; 18,832 respectively. * p<0.1, ** p<0.05, *** p<0.01. Source: American Community Survey from 2012 to 2016.

Table 13. The influence of son preference on gender selection in the U.S

	Male child				
	Full Sample	Natives	First-generation Immigrants	G7 Immigrants	SE Asian Immigrants
Panel A: Son preference for at least one son					
Had Second Child					
All girls	-0.005***	-0.005***	-0.001	-0.004	0.005
	(0.001)	(0.002)	(0.003)	(0.011)	(0.005)
Had Third Child					
All girls	0.019***	0.018***	0.023***	0.039***	0.035***
	(0.002)	(0.002)	(0.003)	(0.014)	(0.005)
Panel B: Son preference controlling for gender balancing effects					
Had Third Child					
All girls	-0.003	-0.005***	0.007*	0.032**	0.024***
	(0.002)	(0.002)	(0.004)	(0.016)	(0.006)
Mix	-0.035***	-0.037***	-0.024***	-0.012	-0.017***
	(0.002)	(0.002)	(0.003)	(0.013)	(0.005)

Notes: The dependent variable, male child, indicates whether the next child is a boy. Sample sizes are as described in Table 12. * p<0.1, ** p<0.05, *** p<0.01.
Source: American Community Survey from 2012 to 2016.

Table 14. The influence of son preference on fertility for first-generation immigrants from Asia

	Fertility					
	China	Japan	Korea	Philippines	Vietnam	India
Panel A: Son preference for at least one son						
Had Second Child [(a)]						
All girls	0.043***	-0.036	0.016	0.013	0.034**	0.029***
	(0.013)	(0.027)	(0.018)	(0.014)	(0.017)	(0.009)
Had Third Child [(b)]						
All girls	0.079***	0.002	0.027	0.043**	0.094***	0.037***
	(0.017)	(0.039)	(0.025)	(0.020)	(0.023)	(0.008)
Panel B: Son preference controlling for gender balancing effects						
Had Third Child						
All girls	0.064***	-0.040	-0.003	0.029	0.069**	0.027***
	(0.019)	(0.045)	(0.030)	(0.024)	(0.027)	(0.010)
Mix	-0.023	-0.068*	-0.046*	-0.021	-0.040*	-0.016**
	(0.015)	(0.037)	(0.024)	(0.020)	(0.022)	(0.008)

Notes: The samples include families who migrated from those Asian countries. The dependent variable, fertility, indicates having another child within six years of previous child's birth. (a). "Had second child" represents the sample of families with at least one child and may have additional children later. The unit of observations is the household, and the sample size for each group is 5,126; 1,364; 2,671; 4,601; 3,099; 11,282 respectively. (b) "Had third child' represents the sample of families with at least two children and may have additional children later. The sample size for each group is 2,620; 775; 1,549; 2,505; 5,655 respectively. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.
Source: American Community Survey from 2012 to 2016.

Table 15. The influence of son preference on gender selection for first-generation immigrants from Asia

	Male Child					
	China	Japan	Korea	Philippines	Vietnam	India
Panel A: Son preference for at least one son						
Had Second Child						
All girls	0.009	-0.024	0.012	-0.000	-0.000	0.011
	(0.012)	(0.025)	(0.018)	(0.013)	(0.017)	(0.008)
Had Third Child						
All girls	0.065***	0.003	0.014	0.016	0.056***	0.029***
	(0.014)	(0.031)	(0.018)	(0.016)	(0.019)	(0.007)
Panel B: Son preference controlling for gender balancing effects						
Had Third Child						
All girls	0.061***	-0.031	0.012	0.013	0.042*	0.021**
	(0.015)	(0.037)	(0.022)	(0.019)	(0.023)	(0.008)
Mix	-0.006	-0.054*	-0.002	-0.005	-0.022	-0.013**
	(0.011)	(0.030)	(0.018)	(0.016)	(0.018)	(0.006)

Notes: The samples include families who migrated from those Asian countries. Sample sizes are as described in Table 12. * p<0.1, ** p<0.05, *** p<0.01.
Source: American Community Survey from 2012 to 2016.

Table 16. The influence of son preference on fertility for higher order generation immigrants from Asia

Higher order generation immigrants	Fertility					
	SE Asian	China	Korea	Philippines	Vietnam	India
Panel A: Son preference for at least one son						
Had Second Child [a]						
All girls	0.010	0.015	-0.002	0.005	0.018	0.035
	(0.010)	(0.021)	(0.030)	(0.019)	(0.042)	(0.027)
Had Third Child [b]						
All girls	0.046***	0.036	0.105**	0.061**	0.073	0.088**
	(0.014)	(0.029)	(0.042)	(0.027)	(0.066)	(0.035)
Panel B: Son preference controlling for gender balancing effects						
Had Third Child						
All girls	0.018	-0.007	0.046	0.049	0.005	0.109***
	(0.016)	(0.035)	(0.052)	(0.032)	(0.077)	(0.040)
Mix	-0.045***	-0.068**	-0.090**	-0.018	-0.098*	0.032
	(0.013)	(0.029)	(0.041)	(0.026)	(0.057)	(0.033)

Notes: The samples contain families who are descendants of immigrants of those Asian countries. The dependent variable, fertility, indicates having another child within six years of previous child's birth. (a). "Had second child" represents the sample of families with at least one child and may have additional children later. The unit of observations is the household, and the sample size for each group is 9,491; 2,067; 1,078; 2,724; 574; 1,238 respectively. (b) "Had third child' represents the sample of families with at least two children and may have additional children later. The sample size for each group is 5,520; 1,154; 626; 1,609; 313; 720 respectively. * p<0.1, ** p<0.05, *** p<0.01.
Source: American Community Survey from 2012 to 2016.

Table 17. The influence of son preference on gender selection for higher order generation immigrants from Asia

Higher order generation immigrants	Male Child					
	SE Asian	China	Korea	Philippines	Vietnam	India
Panel A: Son preference for at least one son						
Had Second Child						
All girls	-0.003	0.007	-0.001	-0.003	0.051	-0.029
	(0.009)	(0.020)	(0.028)	(0.018)	(0.040)	(0.025)
Had Third Child						
All girls	0.019*	0.028	0.074**	0.009	-0.031	0.058**
	(0.011)	(0.025)	(0.033)	(0.021)	(0.051)	(0.027)
Panel B: Son preference controlling for gender balancing effects						
Had Third Child						
All girls	-0.007	0.002	0.023	-0.020	-0.071	0.076***
	(0.013)	(0.029)	(0.041)	(0.025)	(0.062)	(0.028)
Mix	-0.041***	-0.041*	-0.078**	-0.046**	-0.057	0.028
	(0.011)	(0.023)	(0.033)	(0.021)	(0.051)	(0.021)

Notes: The samples contain families who are descendants of immigrants of those Asian countries. Sample sizes are as described in Table 16. * p<0.1, ** p<0.05, *** p<0.01.
Source: American Community Survey from 2012 to 2016.

Table 18. The influence of son preference on fertility for first-generation immigrants from Asia and arrived at the age of 9 above

First-generation immigrants arrived at age 9+	Fertility					
	SE Asian	China	Korea	Philippines	Vietnam	India
Had Second Child [(a)]						
All girls	-0.019	0.001	0.006	0.009	0.049	-0.020
	(0.015)	(0.050)	(0.033)	(0.039)	(0.036)	(0.044)
Arrived at age 9+	-0.096***	-0.106***	-0.082***	-0.055*	-0.038	-0.060*
	(0.011)	(0.037)	(0.030)	(0.029)	(0.031)	(0.032)
All girls*arrived at age 9+	0.048***	0.045	0.011	0.004	-0.020	0.051
	(0.016)	(0.052)	(0.040)	(0.042)	(0.041)	(0.045)
Had Third Child [(b)]						
All girls	-0.029	-0.094	-0.001	0.014	-0.018	-0.122*
	(0.024)	(0.077)	(0.057)	(0.066)	(0.058)	(0.063)
Arrived at age 9+	-0.153***	-0.177***	-0.077*	-0.088*	-0.106**	-0.182***
	(0.018)	(0.062)	(0.046)	(0.050)	(0.048)	(0.049)
All girls*arrived at age 9+	0.069***	0.171**	-0.003	0.017	0.113*	0.156**
	(0.026)	(0.080)	(0.067)	(0.071)	(0.066)	(0.064)
Mix	-0.093***	-0.155**	-0.081*	-0.045	-0.113**	-0.092
	(0.021)	(0.070)	(0.045)	(0.056)	(0.050)	(0.058)
Mix*arrived at age 9+	0.075***	0.144**	0.051	0.028	0.094*	0.081
	(0.022)	(0.071)	(0.053)	(0.060)	(0.056)	(0.059)

Notes: The dependent variable, fertility, indicates having another child within six years of previous child's birth. (a). "Had second child" represents the sample of families with at least one child and may have additional children later. The unit of observations is the household, and the sample size for each group is 34,592; 5,126; 2,671; 4,601; 3,099; 11,282respectively. (b) "Had third child' represents the sample of families with at least two children and may have additional children later. The sample size for each group is 18,832; 2,620; 775; 1,549; 2,505; 1,908; 5,655 respectively. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.
Source: American Community Survey from 2012 to 2016.

Table 19. The influence of son preference on gender selection for first-generation immigrants from Asia and arrived at the age of 9 above

First-generation immigrants arrived at age 9+	Male Child					
	SE Asian	China	Korea	Philippines	Vietnam	India
Had Second Child						
All girls	-0.044***	-0.076	-0.019	-0.029	-0.007	-0.040
	(0.014)	(0.047)	(0.034)	(0.036)	(0.037)	(0.041)
Arrived at age 9+	-0.052***	-0.049	-0.083***	-0.024	-0.002	-0.039
	(0.011)	(0.036)	(0.029)	(0.028)	(0.031)	(0.030)
All girls*arrived at age 9+	0.055***	0.091*	0.042	0.033	0.008	0.052
	(0.015)	(0.049)	(0.040)	(0.038)	(0.041)	(0.042)
Had Third Child						
All girls	0.015	-0.067	-0.008	0.091	0.038	0.045
	(0.019)	(0.057)	(0.039)	(0.056)	(0.047)	(0.054)
Arrived at age 9+	-0.060***	-0.094**	-0.010	-0.018	-0.028	-0.055
	(0.014)	(0.048)	(0.034)	(0.040)	(0.037)	(0.034)
All girls*arrived at age 9+	0.011	0.140**	0.030	-0.095	0.006	-0.026
	(0.020)	(0.059)	(0.047)	(0.060)	(0.054)	(0.054)
Mix	-0.045***	-0.103**	-0.008	0.009	-0.040	-0.051
	(0.016)	(0.051)	(0.033)	(0.045)	(0.039)	(0.039)
Mix*arrived at age 9+	0.034**	0.104**	0.010	-0.016	0.024	0.041
	(0.017)	(0.052)	(0.039)	(0.048)	(0.044)	(0.040)

Notes: Sample sizes are as described in Table 8. * p<0.1, ** p<0.05, *** p<0.01.
Source: American Community Survey from 2012 to 2016.

Table 20. Comparison on migration of immigrants' spouse

Spouse's ethnic group	First-generation immigrants					Higher-order generation immigrants			
	All immigrants	SE Asian	China	Vietnam	India	SE Asian	China	Vietnam	India
First-generation immigrants									
China	0.82	0.80	0.72	0.02	0.01	0.06	0.05	0.00	0.00
Vietnam	0.82	0.81	0.02	0.75	0.01	0.07	0.02	0.04	0.00
India	0.95	0.93	0.00	0.00	0.92	0.02	0.00	0.00	0.02
Higher-order generation immigrants									
China	0.21	0.18	0.07	0.03	0.00	0.32	0.22	0.01	0.01
Vietnam	0.28	0.27	0.01	0.19	0.00	0.27	0.04	0.17	0.01
India	0.38	0.30	0.00	0.00	0.23	0.28	0.01	0.00	0.26

Note: The entry represents the proportion of the race of first-generation or higher-order generation immigrants' husband. For example, the first entry, 0.82, indicates that 82% of First-generation Chinese immigrants (mothers) get married with first-generation immigrants from abroad of the U.S.

Source: American Community Survey from 2012 to 2016.

Table 21. Cox hazard regressions for families' fertility decisions in the U.S.

	Full Sample	Natives	First-generation Immigrants	G7 Immigrants	SE Asian Immigrants
Had Second Child [(a)]					
All girls	-0.007***	-0.009***	0.003	-0.009	0.009
	(0.002)	(0.002)	(0.004)	(0.016)	(0.008)
Had Third Child [(b)]					
All girls	0.003	0.001	0.016	-0.012	0.035
	(0.005)	(0.005)	(0.010)	(0.041)	(0.022)
Mix	-0.022***	-0.019***	-0.031***	0.025	-0.007
	(0.004)	(0.005)	(0.009)	(0.036)	(0.021)

Notes: The dependent variable is the duration between the first and second child as well as the second and the third child. (a). The sample size for each group is 686,575; 539,104; 147,471; 10,044; 44,570 respectively. (b) The sample size for each group is 226,162; 177,083; 49,079; 2,743; 9,678 respectively. * p<0.1, ** p<0.05, *** p<0.01.
Source: American Community Survey from 2012 to 2016.

Table 22. The effects of son preference on mother's marital status

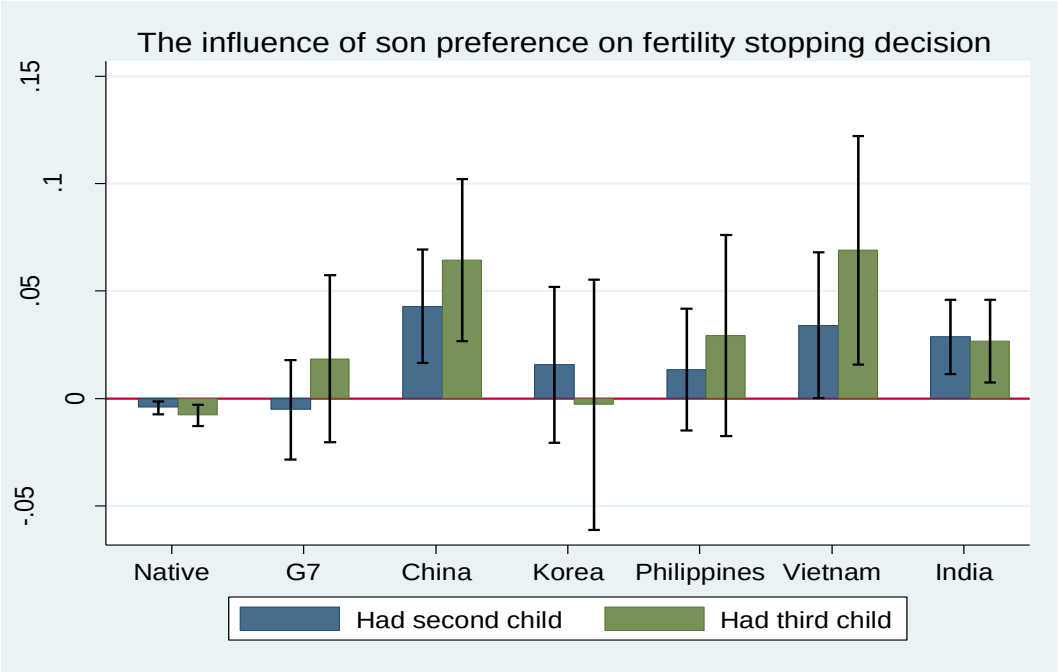
	Full Sample	**Natives**	**First-generation Immigrants**	**G7 Immigrants**	**SE Asian Immigrants**
All girls	0.005***	0.006***	0.002**	0.003	0.002
	(0.000)	(0.001)	(0.001)	(0.003)	(0.001)
Mother's education high school and above	-0.051***	-0.090***	-0.026***	-0.017	-0.036***
	(0.001)	(0.002)	(0.001)	(0.011)	(0.004)
Mother's education bachelor's degree and above	-0.020***	-0.020***	-0.015***	-0.011***	-0.017***
	(0.000)	(0.000)	(0.001)	(0.003)	(0.002)
Mother's age at marriage	0.003***	0.004***	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Family income	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Father's education high school and above	-0.030***	-0.038***	-0.025***	-0.027***	-0.033***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.003)
Father's education bachelor's degree and above	-0.000	0.004***	-0.010***	0.005**	-0.018***
	(0.000)	(0.000)	(0.001)	(0.003)	(0.002)
Mother's age	-0.013***	-0.013***	-0.010***	-0.007***	-0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of children	-0.021***	-0.022***	-0.018***	-0.008***	-0.012***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Number of observations	1,112,703	874,581	238,122	15,920	78,178

Notes: The dependent variable is a dummy equal to 1 if the mother with at least one child is divorced, separated, or never married. * p<0.1, ** p<0.05, *** p<0.01.

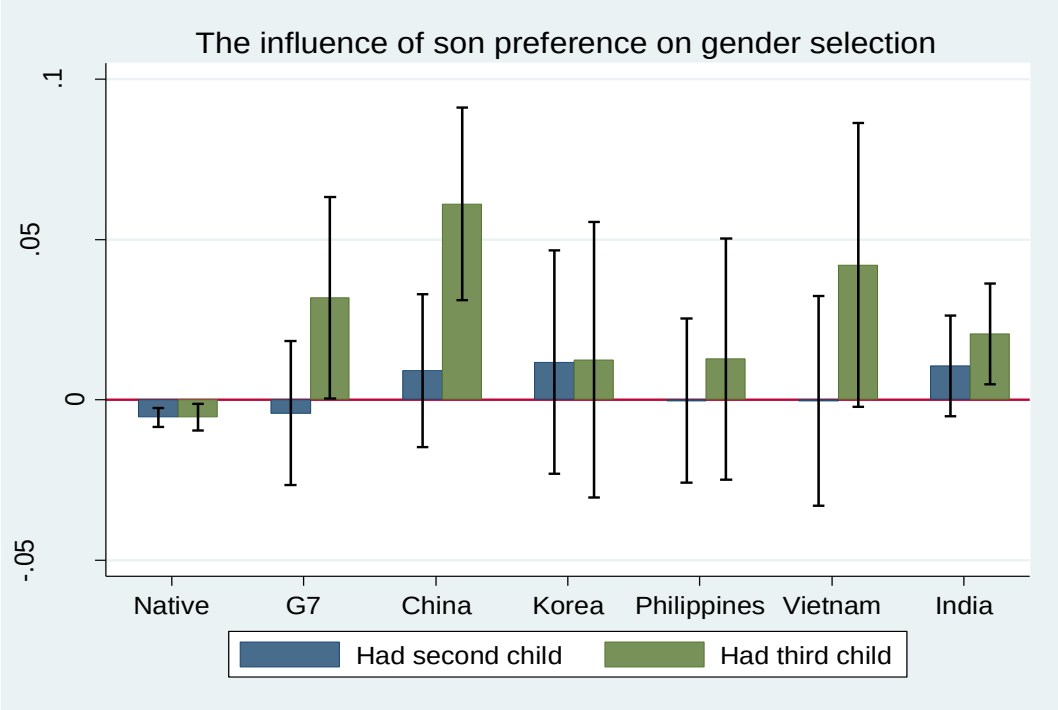
Source: American Community Survey from 2012 to 2016.

Figure 16. The influence of son preference on fertility in the U.S. with confidence intervals

Panel A The influence of son preference on fertility stopping decision



Panel B The influence of son preference on gender selection



Note: Confidence intervel is 95%. Regressions include all covariates described in equation 2, especially controlling for mix.

Figure 17. Son preference by age of migration for families with at least one child

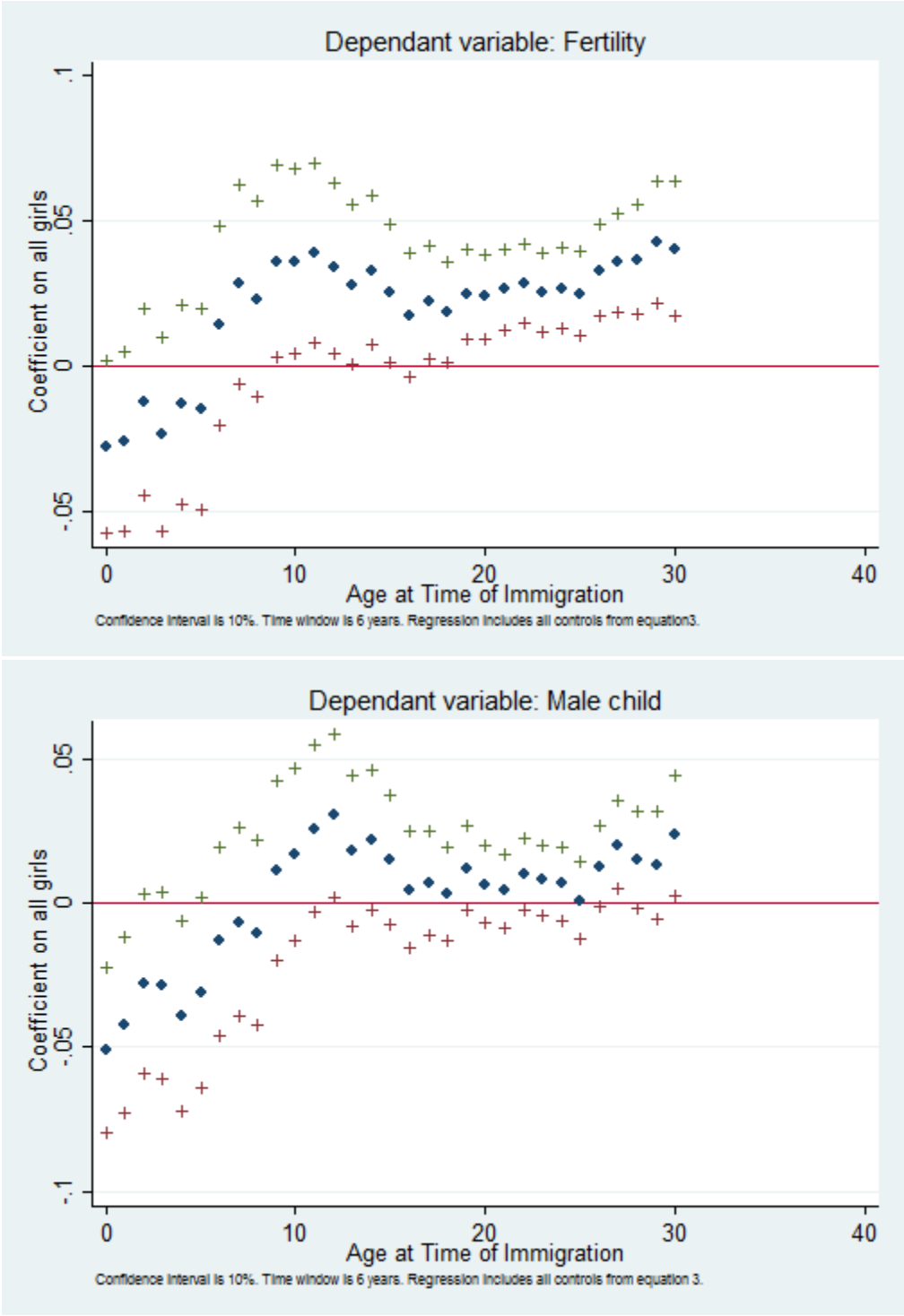


Figure 18. Son preference by age of migration for families with at least two children

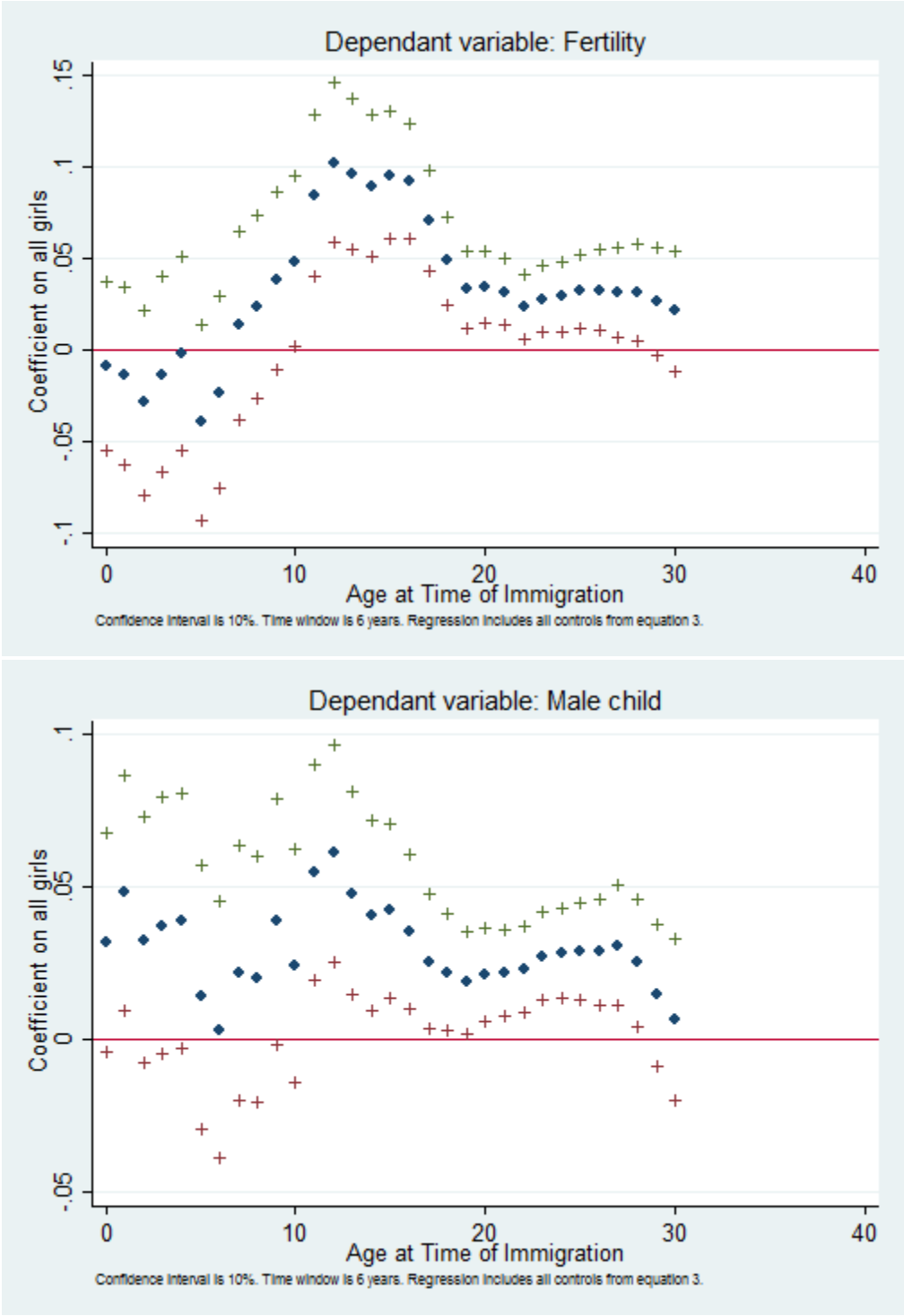


Figure 19. Son preference by years since immigration to the U.S. for families with at least one child

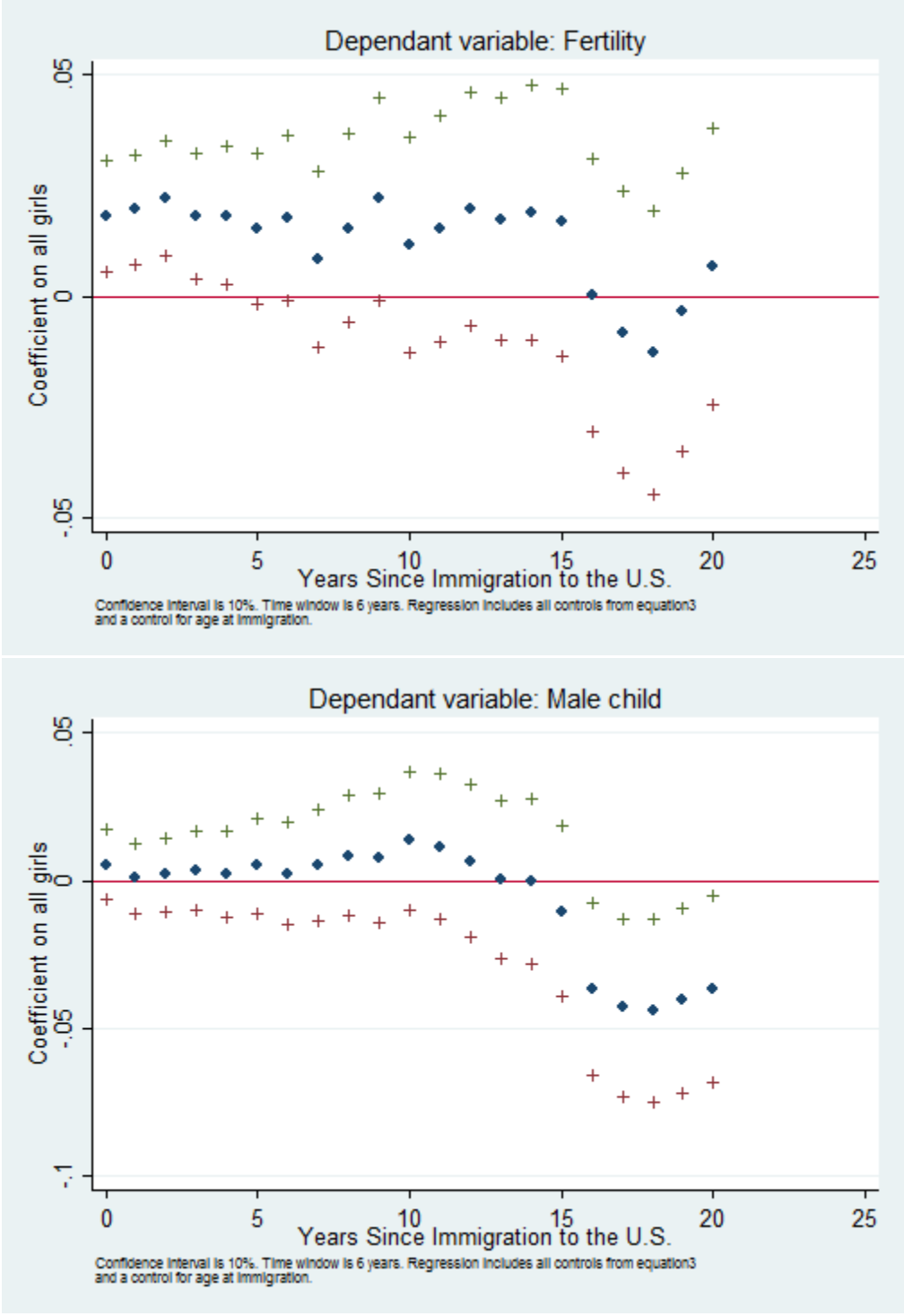
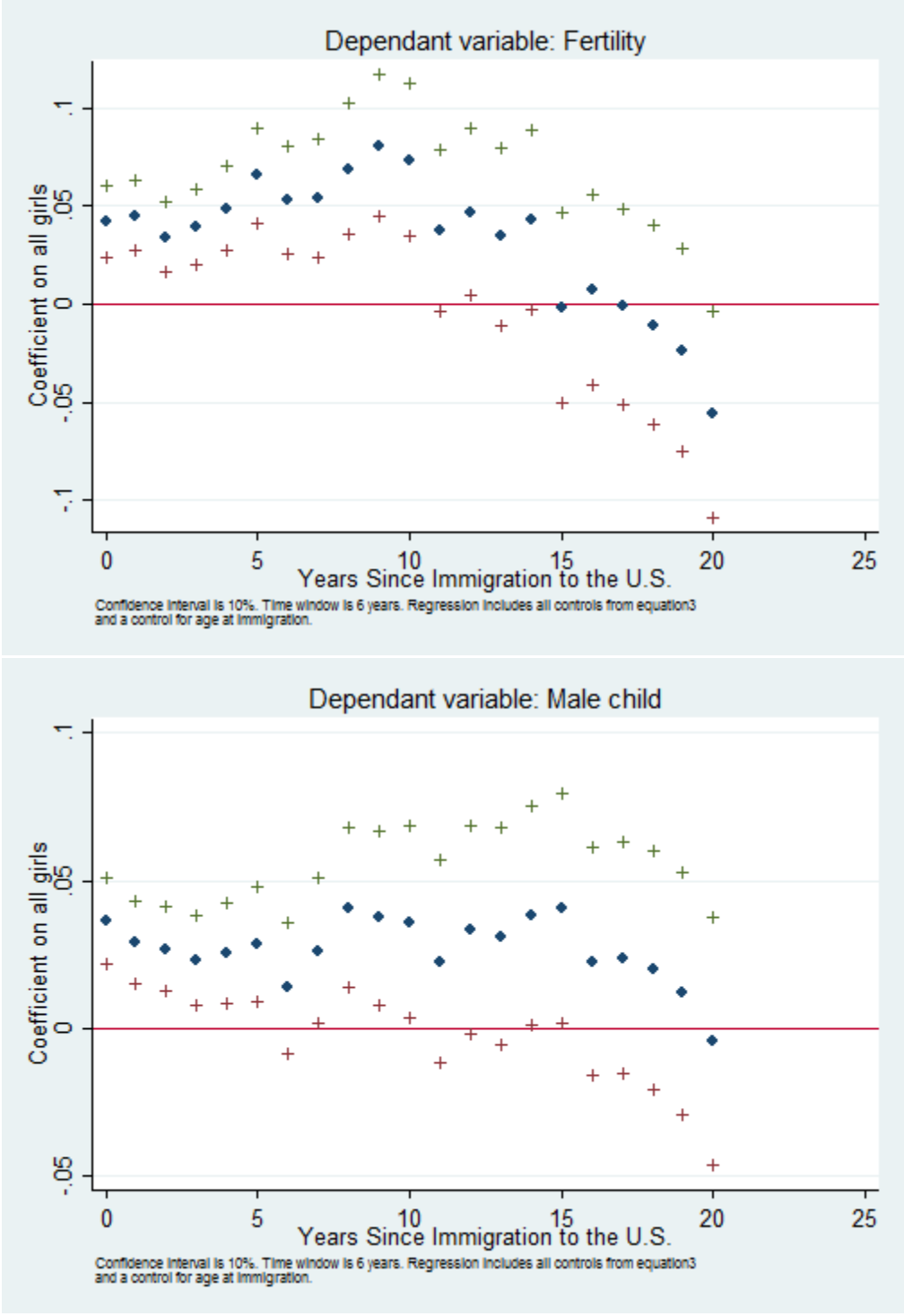


Figure 20. Son preference by years since immigration to the U.S. for families with at least two children



References

Blau, Francine D., et al. "The transmission of women's fertility, human capital, and work orientation across immigrant generations." *Journal of Population Economics* 26.2 (2013): 405-435.

Bleakley, Hoyt, and Aimee Chin. "Age at arrival, English proficiency, and social assimilation among US immigrants." *American Economic Journal: Applied Economics* 2.1 (2010): 165-92.

Abrevaya, Jason. "Are there missing girls in the United States? Evidence from birth data." *American Economic Journal: Applied Economics* 1.2 (2009): 1-34.

Almond, Douglas, Lena Edlund, and Kevin Milligan. "Son preference and the persistence of culture: evidence from South and East Asian immigrants to Canada." *Population and Development Review* 39.1 (2013): 75-95.

Haughton, Jonathan, and Dominique Haughton. "Are simple tests of son preference useful? An evaluation using data from Vietnam." *Journal of Population Economics* 11.4 (1998): 495-516.

Dahl, Gordon B., and Enrico Moretti. "The demand for sons." *The Review of Economic Studies* 75.4 (2008): 1085-1120.

Aly, Hassan Y., and Michael P. Shields. "Son preference and contraception in Egypt." *Economic Development and Cultural Change* 39.2 (1991): 353-370.

Gangadharan, Lata, and Pushkar Maitra. "Testing for son preference in South Africa." *Journal of African Economies*12.3 (2003): 371-416.

Das Gupta, Monica, et al. "Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea." *The Journal of Development Studies* 40.2 (2003): 153-187.

Arnold, Fred, and Liu Zhaoxiang. "Sex preference, fertility, and family planning in China." *Population and Development Review* (1986): 221-246.

Edlund, Lena. "Son preference, sex ratios, and marriage patterns." *Journal of Political Economy* 107.6 (1999): 1275-1304.

Gupta, Monica Das. "Selective discrimination against female children in rural Punjab, India." *Population and development review* (1987): 77-100.

Jayachandran, Seema, and Rohini Pande. "Why are Indian children so short? The role of birth order and son preference." *American Economic Review* 107.9 (2017): 2600-2629.

Kureishi, Wataru, and Midori Wakabayashi. "Son preference in Japan." *Journal of Population Economics* 24.3 (2011): 873-893.

Dubuc, Sylvie, and David Coleman. "An Increase in the Sex Ratio of Births to India-born Mothers in England and Wales: Evidence for Sex-Selective Abortion." *Population and Development Review* 33.2 (2007): 383-400.

Choi, Eleanor Jawon, and Jisoo Hwang. "Child gender and parental inputs: No more son preference in Korea?." *American Economic Review* 105.5 (2015): 638-43.

Blau, Francine D., et al. *Is There Still Son Preference in the United States?*. No. w23816. National Bureau of Economic Research, 2017.

Bleakley, Hoyt, and Aimee Chin. "Language skills and earnings: Evidence from childhood immigrants." *Review of Economics and Statistics* 86.2 (2004): 481-496.