2015

AN ECONOMIC ANALYSIS OF FOOD SAFETY AND PUBLIC HEALTH: MILK PASTEURIZATION IN THE UNITED STATES

Huiqiang Wang
University of Rhode Island, hughwangri@gmail.com

AN ECONOMIC ANALYSIS OF FOOD SAFETY AND PUBLIC HEALTH: MILK PASTEURIZATION IN THE UNITED STATES

BY

HUIQIANG WANG

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ENVIRONMENTAL AND NATURAL RESOURCE ECONOMICS

UNIVERSITY OF RHODE ISLAND

2015

DOCTOR OF PHILOSOPHY DISSERTATION

OF

HUIQIANG WANG

APPROVED:

Dissertation Committee:

Major Professor __Thomas W. Sproul_____

_____ Corey Lang _____

_____John Burkett_____

_____Nasser H. Zawia_____
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2015

ABSTRACT

My dissertation is a comprehensive economic history study to the public health impacts of milk pasteurization in the United States. It has four major focuses which are included into four chapters. Chapter I is a case study to the public health impact of Chicago's pasteurization ordinance. This study sets up the causal relationship between milk pasteurization and health outcomes. Chapter II extends a new econometric tool, the synthetic control methods, from a single unit to multiple treated units. This chapter also measures the impacts of pasteurization ordinances in a group of cities. Chapter III is written more from an econometric perspective. It concerns how the synthetic control method can be transformed into a linear regression based model, which has more potential for empirical policy evaluations. Chapter 4 takes an alternative view to milk pasteurization. It discusses how the extent of pasteurization could make difference to public health. It also compares estimations of regular least square model and robust panel data model.

Using Chicago's 1916 pasteurization ordinance as a comparative case study, the first chapter focuses on how to measure the health impacts of food safety interventions. Empirical evidence suggests there was a clear causality relation between milk pasteurization and variations in the health outcomes of interest in Chicago. Thus, I applied the non-parametric synthetic control approach to capture causal health effects of this ordinance. The results suggest that the effect of this policy intervention was more pronounced in Chicago than in its 20 comparison cities, so I conclude that Chicago's 1916 pasteurization ordinance had positive health effects.

The second chapter examines causal health effect of mandatory city pasteurization ordinances in the United States. I apply the synthetic control methods to multiple treated units (MTSCM). Results indicate noticeable health benefits are observed in some cities but not all. For inferences, non-parametric rank-sum tests are preferred because of non-normal outcomes in the control group. This study also suggests regression based Difference-in-Difference (DD) models lead to different results than SCM, since SCM reveals more information like unit-varying and time-varying treatment effect.

The third chapter aims to provide a robustness test for major conclusions obtained from prior chapters, *e.g*. the effect of Chicago's 1916 milk pasteurization ordinances. Using the synthetic control methods (SCM), I found a significant treatment effect. To verify SCM results, I use a linear regression based cross-sectional time series model (CTM) to re-estimate this intervention. CTM results confirm major findings in my prior SCM studies. In addition, I use the 1989 California cigarette sales tax as an "out-of-sample" robustness check for CTM. Again, CTM results are similarly significant as SCM.

The last chapter measures health impacts of variations of extent of pasteurization. Empirically, I choose the Fixed-Effects model to control unobserved intra-city variations. With respect to influential observations, I use robust estimators to validate least squares estimations. Compared with OLS estimate, robust estimates of the coefficients are smaller in absolute value. But their standard errors are even lower. In sum, my FE regressions also support the positive health effect of pasteurization.

ACKNOWLEDGMENTS

This study is finished with generous help from many faculty and graduate students at University of Rhode Island (URI). I would like to thank all economists and non-economists who have contributed to my dissertation. Special thanks to my major professor Thomas W. Sproul who has helped me to finish my graduate program and my dissertation. I also owe a debt of gratitude to my core committee member and a mentor at URI, Professor John Burkett. He provided a lot of advice for polishing my dissertation and worked together with Professor Sproul to finish the final revisions. I am also grateful to Professor Corey Lang in my department and Professor Tong Yu from Business School. As committee members, they contributed insightful guidance and advice to this study. Thanks also to Professor and Chair James J. Opaluch, Professor Hirotsugu Uchida, Professor Emi Uchida and Ms. Judith Palmer in my department. I am also grateful to Professor Cathy Roheim, who was my first mentor at URI and now an agricultural economics professor and department chair at University of Idaho. Special thanks go to two professors from Department of Marine Affairs at URI. I am grateful to a knowledgeable ocean policy expert, Professor Lawrence Juda and his help for my dissertation proposal writing, Last but not least, Professor Trace Dalton, I owe a favor to her. She was the chair in my dissertation defense committee.

In addition, I appreciate all supports from graduate students in my department. Special thanks to my family members and other friends in Rhode Island.

PREFACE

This dissertation is written in a manuscript format. The first chapter follows the format of Journal of Economic history. The second chapter follows the guideline of American Journal of Agricultural Economics. The third and fourth chapters are similarly formatted as the second chapter. Captions of Tables and Figure in Chapter I are formatted as Journal of Economic History, which are written in the capitalized letters. For clarity and consistency, Tables and Figures are number as Chapter #- Table # (*e.g*. Table 1-2) and Chapter #- Figure # (*e.g*. Figure 3-5). In each chapter, only Table # and Figure # are reported. Footnotes in Chapter I are also made according to the format of Journal of Economic history. Footnotes in other chapters are in one format.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER I.

MILK AND CHCICAGO'S MORTALITY TRANSITION IN THE EARLY 1900S: A CASE STUDY OF THE 1916 PASTEURIZATION ORDINANCE

Huiqiang Wang

Department of Environmental and Natural Resource Economics,

University of Rhode Island, Kingston, RI, USA 02881

(In the format of Journal of Economic History)

July 2015

ABSTRACT

Using Chicago's 1916 pasteurization ordinance as a comparative case study, this paper focuses on how to measure the health impacts of food safety interventions. Empirical evidence suggests there was a clear causality relation between milk pasteurization and variations in the health outcomes of interest in Chicago. Thus, I applied the non-parametric synthetic control approach to capture causal health effects of this ordinance. The results suggest that the effect of this policy intervention was more pronounced in Chicago than in its 20 comparison cities, so I conclude that Chicago's 1916 pasteurization ordinance had positive health effects.

KEYWORDS: Public Health; Pasteurization; Chicago; Synthetic Control

1. INTRODUCTION

The availability of higher-quality milk has been regarded as a critical factor in the decline of infant and early childhood mortality. With industrialization and the increased popularity of bottle feeding, milk-borne diseases and their effects on childhood mortality became a worldwide social problem in the late 1800s.[1] As early as the 1910s, health experts informed the public that contaminated milk had become the most common source of food-borne contagious illness.[2] Milk-borne epidemics include typhoid fever, influenza, diphtheria, non-lung tuberculosis, and diarrhea. Early experiments found that heating milk could reduce the number of bacteria and preserve quality, and as a result, pasteurization was introduced in the late 1800s and started to become widespread in the US. The health effects of pasteurization were remarkable. Evidence in the early 1900s indicated that milk-borne illness mortality was lower if children were fed pasteurized milk. Very few children died of summer diarrhea, which was a leading cause of death in infants and young children at that time. Medical professionals recommended pasteurization as a feasible way to keep milk clean and pure and determined it an "essential safeguard" to milk quality.[3]

In the US, pasteurization became a controversial topic. First, the most active opposition came from dairy farmers and dairy organizations. They fiercely fought bovine tuberculin tests and compulsory pasteurization. [4] Second, consumers worried

[1] Beaver, "Population, Infant Mortality, and Milk"; Selitzer, "The Dairy Industry in America," p. 111–135; Atkins, "Mother's Milk and Infant Death in Britain;" Wolf, "Low Breastfeeding Rates and Public Health in the United States;" Lee, 2007 "Infant Mortality Decline in the Late 19th and Early 20th Centuries."
[2] Robertson, "Annual Report of the Department of Health of the City of Chicago," p. xiii.
[3] Winslow, "Man and Epidemics", Chapter 5, "Milk Supply," p. 115–126.
[4] Olmstead and Rhode, 2007, "Not on my Farm! Resistance to Bovine Tuberculosis Eradication in the United States," p. 782.

that pasteurization could change milk quality and lead to nutritional loss.[5] Resistance to pasteurization was also observed in Canada and the UK.[6] As a result, regulations to pasteurize milk for home consumption lagged behind commercial applications. In the US, city-level ordinances began in the 1910s, state-level regulation did not follow until the 1940s, and the federal ban of unpasteurized milk for interstate trade was not enacted until 1987. As of 2013, public attitudes toward this technology were still split. Twenty states still allow the sale of raw milk within state borders, while 30 states ban it. Similar to the case of bovine tuberculosis eradication in the US before 1940, milk pasteurization was an icon as to how safer milk could positively affect public health. [7]

This paper focuses on Chicago as a case study because in 1908 it was the first US city to adopt a compulsory pasteurization ordinance. However, it was subsequently blocked by the State of Illinois after dairy farmers rallied the Illinois legislature to oppose this municipal policy.[8] It wasn't until 1916 that Chicago fully implemented its pasteurization ordinance. [9] Historical facts tell us the health consequences were obvious, but they have never been quantified rigorously due to some empirical challenges: outbreaks of infectious diseases were largely well controlled at the time and infant and childhood mortality also declined over this period. The historical significance has been well documented in the literature, but quantitative discussions are still rare. Therefore, this study aims to estimate the health benefits of Chicago's 1916 ordinance, which could provide implications to modern food safety policies.

[5] Hall and Trout, "Milk Pasteurization" (early oppositions to the pasteurized milk in the US).
[6] Phillips and French, "State Regulation and the Hazards of Milk, 1900–1939," p. 371–388.
[7] Olmstead and Rhode, 2004, "An Impossible Undertaking: The Eradication of Bovine Tuberculosis in the United States," p. 743.
[8] Olmstead and Rhode, 2007, ibid.
[9] Czaplicki, "Pure Milk is Better than Purified Milk." On state level, the first 100% milk pasteurization laws were made in Colorado and Utah in1947, see Dahlberg and Adams (1950).

Empirically, there are two major challenges in identifying the effects of this ordinance. First, I need to confirm the causal relation between milk pasteurization and health consequences. In other words, I need to make sure pasteurization determined the transition of our dependent variable. Second, I need a proper empirical model to identify the causal effect of our target policy intervention. In this study, the choice of models was constrained by data availability and the need for a proper counterfactual.

The first question is the basic building block of this study. In addition to pasteurization, other public health campaigns occurred in that period. These campaigns are confounding factors for my analysis, and some of them had profound impacts, such as water purification measures. This type of uncertainty could be a potential threat to the validity of our identification strategy. In Chicago, the adoption of pasteurization overlapped with water purification measures in the 1910s. Contaminated water was a major source of some infectious diseases. Quantitative analyses[10] have measured the contribution of water filtration and chlorination to the drop in typhoid fever deaths. In comparison, the role of milk quality in the decline in childhood mortality is less clear. Thus, I will attempt to show whether pasteurization was the true cause of changing childhood mortality in Chicago in the 1910s.

The second issue is what empirical strategy should be used to identify the causal effect of the 1916 policy intervention? Estimation and causal inference for this type of historical policy evaluation is not easy due to the presence of unobserved heterogeneity in the data and identifying appropriate comparison units (the control group). Ordinary Least Squares (OLS) regression cannot control for unobserved

[10] Cutler and Miller, "The Role of Public Health Improvements in Health Advance: The Twentieth-century United States;" Ferrie and Troesken, "Water and Chicago's Mortality Transition: 1850–1925."

heterogeneity, while Fixed Effects (FE) and Difference-in-Difference (DD) estimators require that unobserved factors remain constant over time for individuals or constant within time periods, the "constant trend assumption."[11] In addition, there is an issue of choosing suitable control cities and weighting them appropriately, [12] because inappropriate comparison units may lead to erroneous conclusions in comparative case studies. Different outcomes of the treated and control sets may reflect disparities in their characteristics,[13] instead of the intended identification of a treatment effect.

Motivated by these concerns, this study proceeds as follows (Figure 1). The first step is to confirm the health impacts of milk pasteurization. In order to be an effective policy intervention, the 1916 ordinance would have to demonstrate a structural change in the trajectory of the health outcome. Also, I will discuss how to choose appropriate dependent variables which could reflect the effect of this ordinance. This step will separate out confounding influences from other factors, largely water purification. Empirical evidence suggest that the pasteurization ordinance made a unique contribution to public health, as it was the leading factor in the decline of diarrhea-related mortality in infant and one-year old children. Further, water treatment had no clear correlation in the decline of early childhood diarrhea cases in Chicago, although it was related to a drop in typhoid fever mortality.

[11] Wagstaff, "Estimating Health Insurance Impacts under Unobserved Heterogeneity," p. 190–191.
[12] King and Zeng, "The Dangers of Extreme Counterfactuals," p. 132.
[13] Abadie, Diamond, and Hainemuller, "Comparative Politics and the Synthetic Control Method", p.1–3.

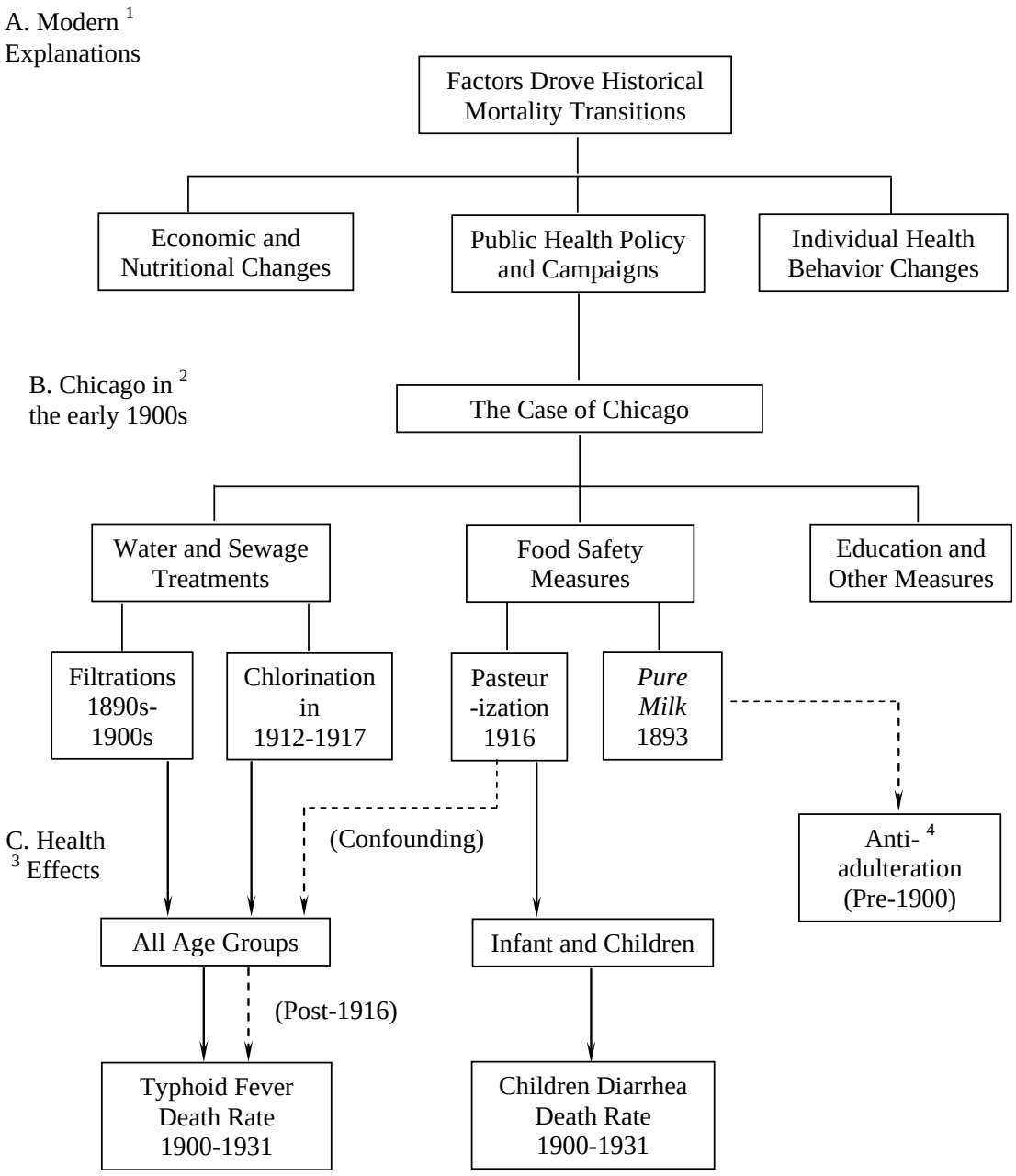


Figure 1- 1DIAGRAM OF KEY FACTORS

Notes: 1. The three major factors are from Cutler and Miller (2005). 2. More details are in Robertson (1919). 3. The mortality rates of typhoid fevers have been used to measure the effect of water treatments, for example Cutler and Miller (2005); Ferrie and Troesken (2008). 4. Early milk laws were adopted to eliminate adulterations and maintain fat and protein contents. These measures worked well in Chicago before 1900, see Alvord (1903).

In the second step, I estimate the causal health impact of the ordinance using the Synthetic Control Method (SCM). SCM was originally designed for case studies and is robust to unobserved heterogeneity over time. The method uses an optimized weighting procedure to get a better counterfactual for estimating the effect of an intervention. Roughly speaking, SCM has the advantages of DD and Propensity Score Matching (PSM) methods over a broader range of data-generating processes.[14] The SCM results show that the effect of Chicago's ordinance was higher than all its comparison units and that the decline in mortality was significant.

The rest of this paper is organized as follows. The second part reviews the health impacts of milk throughout history and the major public health campaign in Chicago in the early 1900s. Next, I analyze empirical evidence about the trend of Chicago's diarrhea and typhoid fever mortality data, aiming to confirm the role of pasteurization. In the following part, I set up the framework of causal effect estimation and discuss constraints of standard methods. Then, I apply SCM to estimate causal health effects of the 1916 ordinance and make inferences. The last section concludes.

[14] Developed in Abadie and Gardeazabal, "The Economic Costs of Conflict;" and Abadie, Diamond, and Hainmueller, "Synthetic Control Method for Comparative Case Studies."

2. BACKGROUND

Milk and Health in History

Historically, infant mortality was regarded as an important indicator of public health and social welfare. In the US and Western Europe, infant mortality has experienced a sharp decline in past 200 years, with gradual declines in the second half of the 1800s, and sharp declines in the1900s. This trend can be explained by a combination of factors, such as economic growth, improved public sanitation and medical provisions, dissemination of childcare techniques and knowledge, and improved food and nutrition,[15] particularly, the contribution of improved availability and quality of cow's milk, which became a popular infant food in the late 19th century.[16] On one hand, the increased availability of milk bolstered infant nutrition while on the other, pasteurization and other technological breakthroughs largely reduced the occurrences of milk-borne disease and related infant mortality.[17]

In the early 1900s, poor quality milk was responsible for hundreds of thousands of US deaths[18] and was the leading factor contributing to the extraordinarily high infant death rate. It became a social problem the late 1800s to the early 1900s. This problem was even more serious in cities, where milk could be transported for one hundred miles or more from outside city boundaries. Without adequate refrigeration,

[15] Beaver, ibid; Preston and Haines, "Fatal Years: Child Mortality in Late Nineteenth-century;" Haines, "Inequality and Childhood Mortality;" Haines, "American Fertility in Transition: New Estimates of Birth Rates in the United States;" Haines, "Inequality and Infant and Childhood Mortality in the United States in the Twentieth Century;" Lee, 1991, "Regional inequalities in infant mortality in Britain, 1861–1971;" McKeown, "The Modern Rise of Population; " Fogel, "Economic Growth, Population Theory;" Millward, and Frances, "Economic Factors in the Decline of Mortality in Late Nineteenth Century Britain."

[16] Beaver, ibid; Meckel, "Save the Babies: American Public Health Reform and the Prevention of Infant Mortality, 1850–1929;" Preston and Haines, ibid; Lee 2007, ibid.

[17] Selitzer, ibid, p. 129–135.

[18] Olmstead and Rhode, 2004, ibid, p. 766.

milk could become contaminated by bacteria overnight.[19] In some cities, the number of bacteria in market milk was higher than that found in sewer water.[20] Early researchers found contaminated milk was a source of many epidemics, including diarrhea, typhoid, cholera, scarlet fever, and other infectious diseases. From 1870 to 1900, diarrhea was responsible for about 50 percent of infant mortality among all infectious diseases in the US. [21]

Responding to increasing demand for safe milk in cities, commercial milk supplies began to be pasteurized around 1890. Compared to other safe milk products, like certified milk, pasteurization is more technically consistent and cost effective. Olmstead and Rhode (2004) explained why "certified dairies" were not able to provide enough protection against bovine tuberculosis. One problem was infrequent and lax dairy and herd inspections. Moreover, certified milk was more expensive and comprised no more than 2 percent of the total milk supply in the market.[22] Later some cities began to adopt mandatory ordinances requiring most milk sold to be pasteurized before sale. In the US, Chicago was the first city to require pasteurization. However, this ordinance was later banned by the State of Illinois. In Canada and Europe, supplying an adequate, safe milk supply was also an important task of urban authorities.[23] In Germany, municipal authorities began to control adulterations and bacterial contamination in milk, and initiated a public milk distribution system

[19] Selitzer, ibid, p. 113–115.
[20] Selitzer, ibid, p. 129–135.
[21] Lee, 2007, ibid, p. 586.
[22] Olmstead and Rhode, 2004, ibid, 742; More Discussions can be found in MacNutt, "The Modern Milk Problem;" Kelly and Clement, "Market Milk;" and Block, "Purity, Economy, and Social Welfare in the Progressive Era Milk Movement."
[23] Beaver ibid; Vögele and Woelk, "Public Health and the Development of Infant Mortality in Germany;" Jenkins "Region, Politics, Pasteurization, and the Naturalizing Myth of Pure Milk in 1920s Saint John, New Brunswick."

beginning in the early 1900s. The health benefits of safe milk were noticeable.[24] In the UK, contamination of market cow's milk was also very serious. The spread of bovine tuberculosis and summer diarrhea resulted largely from unsafe milk and was responsible for infant mortality. [25]

Public Health Campaigns in Chicago

As mentioned previously, Chicago was the first city to adopt a mandatory pasteurization ordinance in the US, but the requirement of full milk pasteurization was not implemented until 1916. Besides milk, water was another key factor in Chicago's transition to lower mortality. Ferrie and Troesken (2008) examined the role of a clean water supply on Chicago's public health in 1850–1925. Their results confirmed that the drop in Chicago's total mortality rate was led by much lower childhood infectious disease deaths in that period. They also noticed the positive effects of early water filtration and chlorination to reduce typhoid fever, which accounted for a 35 percent mortality decrease.

In Chicago, three large-scale water purification projects were completed from 1870 to 1920. The first was a two-mile tunnel in 1870 which extended Chicago's water source from the heavily polluted shorelines of Lake Michigan. The second was a four-mile water intake crib in 1893 and a drainage canal in 1900. The third was water chlorination during the period 1912 to 1917. The 1900 drainage canal changed Chicago's sewage disposal flows, after which the flow of the Chicago River, which was carrying sewage into Lake Michigan, was reversed. This was a critical step in

[24] Vögele and Woelk, ibid, p. 591–594.
[25] Atkins, 2000, "The Pasteurization of England;" Atkins, 2003, ibid.

preventing water-borne diseases.[26] At the same time, Chicago's milk quality control programs were also imposed. Under supervision of the city health department, Chicago's municipal milk quality standards were adopted in 1892.[27] In 1908, Chicago passed its first city-level milk pasteurization ordinance in the US, but it was later banned by the State of Illinois. Finally, on July 22 1916, Chicago issued a full pasteurization ordinance.[28]

In addition, other public health campaigns were conducted by Chicago's Department of Health. According to the Department's report, their efforts included offering courses to the Little Mothers' Clubs (Roberston 1919). Over 8,900 girls received certificates for completing the course. Further, for a short while the Department printed a special publication entitled, *Clean Living Magazine*. To control influenza epidemics, smoking compartments were removed on the city's surface and elevated trains. A contagious disease hospital was built to provide quarantine areas for those inflicted with diphtheria and scarlet fever. Besides its health department, Chicago's Health Association also provided public health education (and similar programs were available in many cities). Door-to-door hygiene campaigns were also supported by health spending at that time (Miller 2008).[29]

[26] Ferrie and Troesken, ibid. p. 2–4.
[27] Alvord and Pearson, "The Milk Supply of Two Hundred Cities and Towns", p. 62–66.
[28] Robertson, ibid. p. xv.
[29] Miller, "Women's Suffrage, Political Responsiveness, and Child Survival in American History," p. 1287–1289.

3. CHICAGO'S MORTALITY TRANSITION

Infant and Childhood Mortality

In the early 1900s, the total mortality rate change was driven by a decline in infant and childhood deaths. Chicago was a good example of this trend. Figure 2 plots the shares of infant (under one year) and childhood (under two years) total mortality. Overall, infant deaths comprised over 20 percent of total mortality in 1900–1910; this increased to approximately 25 percent if one-year old children are included. Infant mortality gradually declined after that time, and was under 10 percent in 1930. The number of infant and childhood deaths is reported in Figure 3. Although their share in total mortality was declining, the number increased before 1910 and the decline was not realized until the latter part of the decade. The peak number of infant deaths was 6,939 in 1913. A similar decline is observed regarding deaths of one-year old children.

As reported earlier, contaminated milk and water were correlated with high occurrences of diarrhea and typhoid fever. Prior studies illustrated how water purification measures helped fight typhoid fever mortality.[30] They found that water filtration and chlorination markedly decreased the typhoid fever mortality rate. But the relationship between water, milk, and diarrhea-related mortality was less clear. At that time, quantitative analyses were rare. Medical and public health studies reported that diarrhea could be either water- or milk-borne. In other words, water could be a confounding factor in our study of the effect of the pasteurization ordinance, if I use overall diarrhea-related mortality as the dependent variable. However, with a closer look at age groups, I found milk-borne diarrhea mortality was more prevalent in children under two years of age. Most nutrition and water intake of children in this age

[30] Cutler and Miller, ibid; Ferrie and Troesken, ibid.

group were from milk; either breast milk or market cow's milk. So the problem of confounding influence would be much lower if I focus on the under one-year old and under two-year old age groups (Figure 1).

Figure 4 plots the numbers of typhoid fever and diarrhea-related deaths in Chicago, and Figure 5 shows their shares in total deaths. The number of diarrhea-related deaths was much higher than those of typhoid fever. Total typhoid fever deaths peaked in 1902 at 819. In that year, diarrhea caused 2,188 deaths. In 1916, diarrhea-related deaths reached their highest level at 3,872, while there were only 130 typhoid fever-related deaths. Overall, typhoid fever accounted for less than 5 percent of total mortality; the share of diarrhea-related mortality was higher. Second, the trends of typhoid fever- and diarrhea-related mortality were different. For typhoid fever, after a spike in 1902, it declined. However, the trend of diarrhea-related deaths was quite different. They remained unchanged from 1900 to 1903, when typhoid fever-related deaths peaked. Then, during 1903–1910 diarrhea-related deaths increased, while typhoid fever-related deaths decreased. With ups and downs, the number of diarrhea-related deaths peaked in 1916, followed by a long-term decline until the 1930s.

Age group analysis is appropriate and quite useful here, since infants and one-year old children are vulnerable to disease due to low-quality milk. Figure 6 shows typhoid fever caused only 0.1 percent of total deaths in the under one-year age group from 1900–1910. Even in the one-year old group, this disease caused less than one percent of deaths in that period. In contrast, this disease caused a much greater share of deaths in the two-year and three-year old and above age groups. One thing that was identical across groups was that the percentage of typhoid fever-related deaths quickly

declined in the observation period 1900–1931. Figure 7 depicts the shares of under one-year and under two-year olds in total typhoid fever mortality. It shows that most typhoid fever-related deaths were from the above two-year old population. In most years, children in the two groups accounted for less than 5 percent of total typhoid fever deaths.

Similarly, Figure 8 shows the share of diarrhea-related deaths in total mortality across age groups. In the under one-year old group, the share of diarrhea-related death was remarkably high, at more than 30 percent of total mortality. This increased to more than 40 percent in 1910, peaking in 1916. This pattern is also observed in other age groups. For older age groups, the share was much lower. The percentage of these deaths in the three-year old and above age group was quite small, at 1.7 percent in 1900. It never reached above 1.5 percent after 1905. Figure 9 illustrates that most diarrhea-related mortality occurred in the under two-year old age group. Infants (under one-year of age) comprised 65 percent of the total diarrhea-related mortality in 1900, increasing to 79 percent in 1911. Similarly, diarrhea-related mortality in the under two-year old group was 82 percent of total mortality in 1900 and 92 percent in 1910. Figures 10 and 11 illustrate the number of diarrhea-related deaths by age. The number of diarrhea-related deaths in both under one-year old and under two-year old groups increased until 1910 (Figure 10). In Figure 11 the increasing trend in the two-year old group is also indicated, but at a lower slope. The case of children aged three years and above is different. Deaths in this group realized a declining trend from 1900 to the middle 1910s.

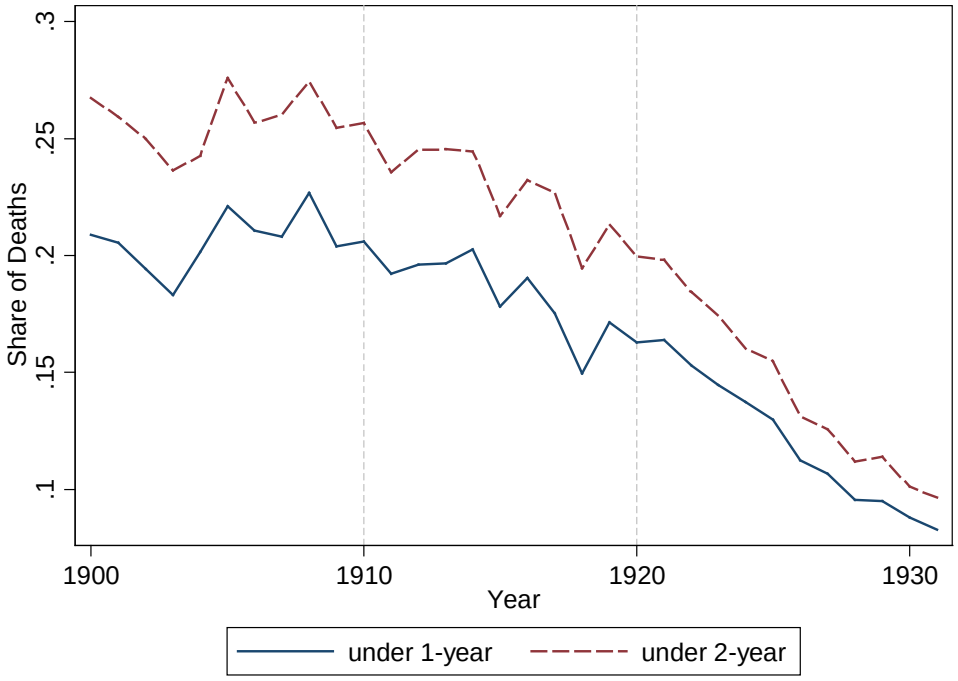


Figure 1- 2 SHARES OF INFANT AND CHILD DEATHS IN TOTAL MORTALITY 1900–1930

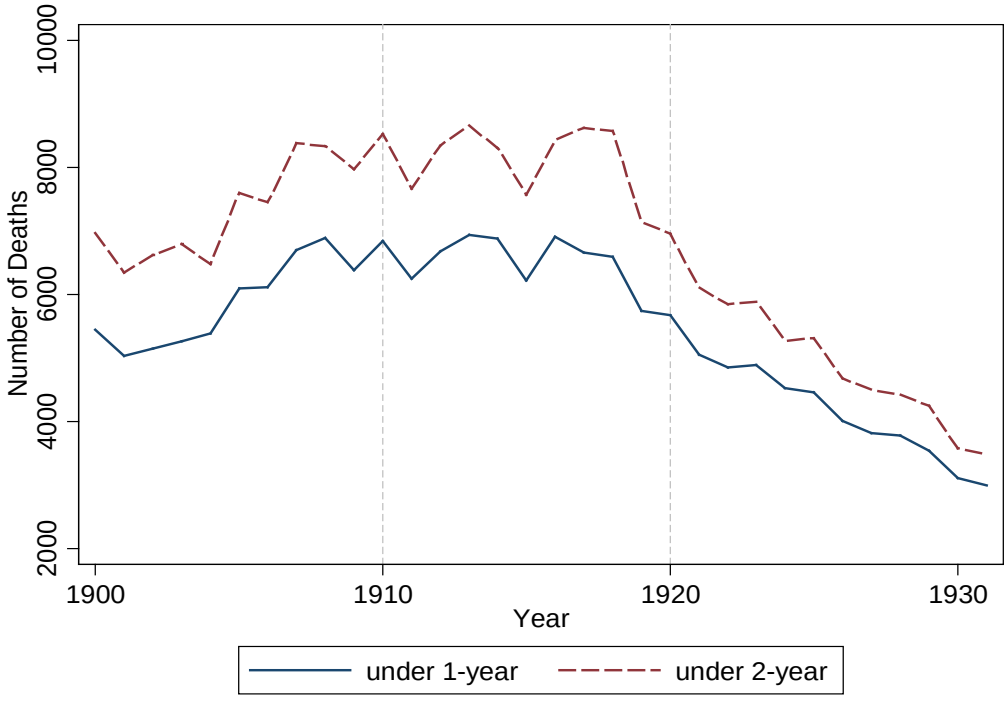


Figure 1- 3 NUMBER OF INFANT AND CHILD DEATHS IN CHICAGO

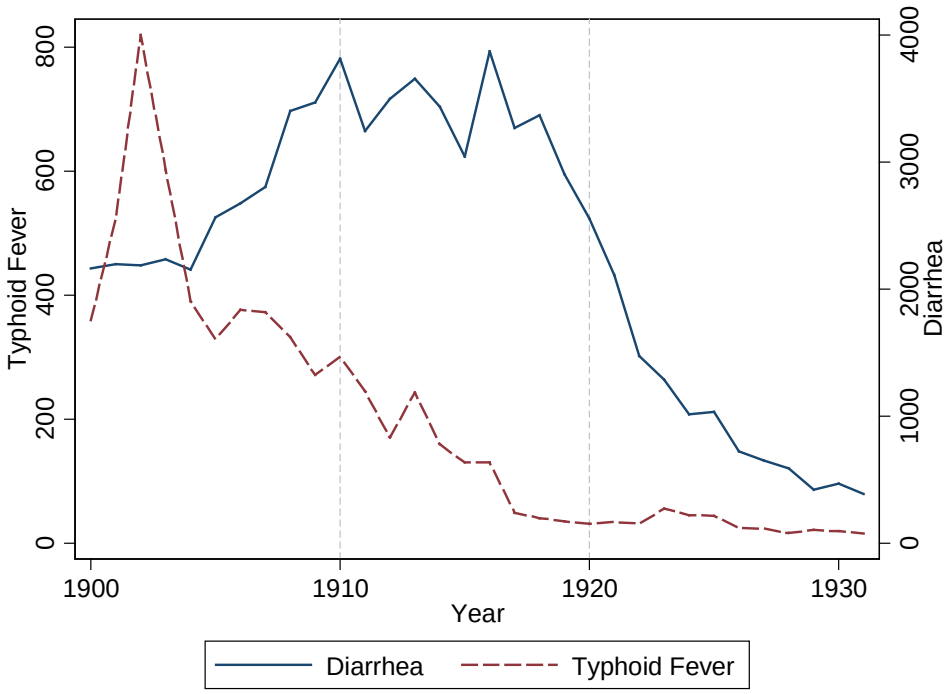


Figure 1- 4 NUMBER OF DIARRHEA- AND TYPHOID FEVER-RELATED DEATHS IN CHICAGO

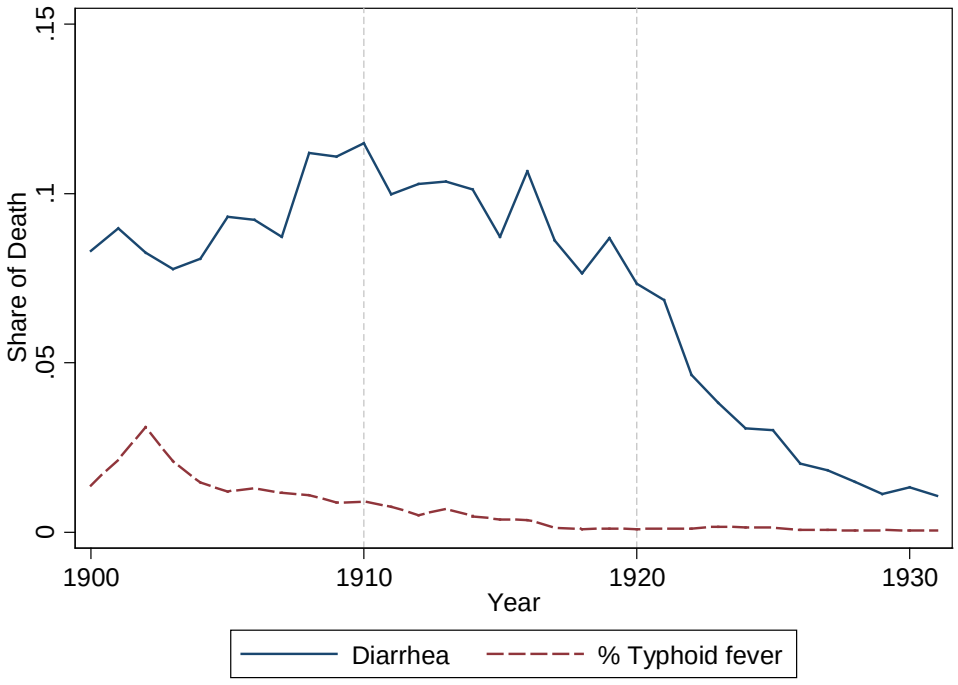


Figure 1- 5 SHARE OF TOTAL DIARRHEA- AND TYPHOID FEVER-RELATED DEATHS

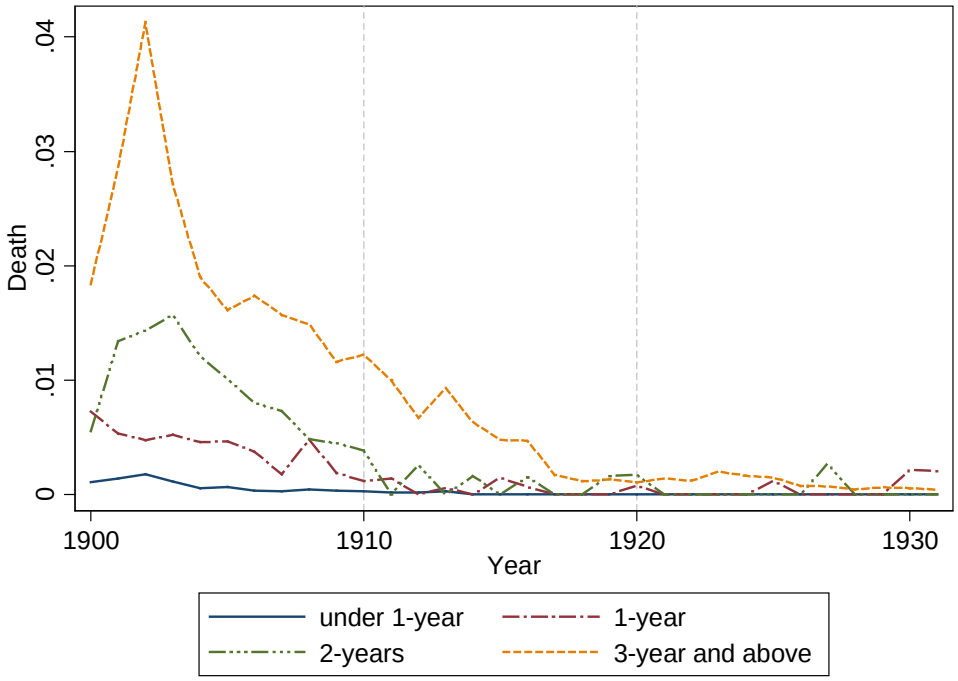


Figure 1- 6 SHARES OF TYPHOID FEVER IN TOTAL MORTALITY BY AGE GROUP

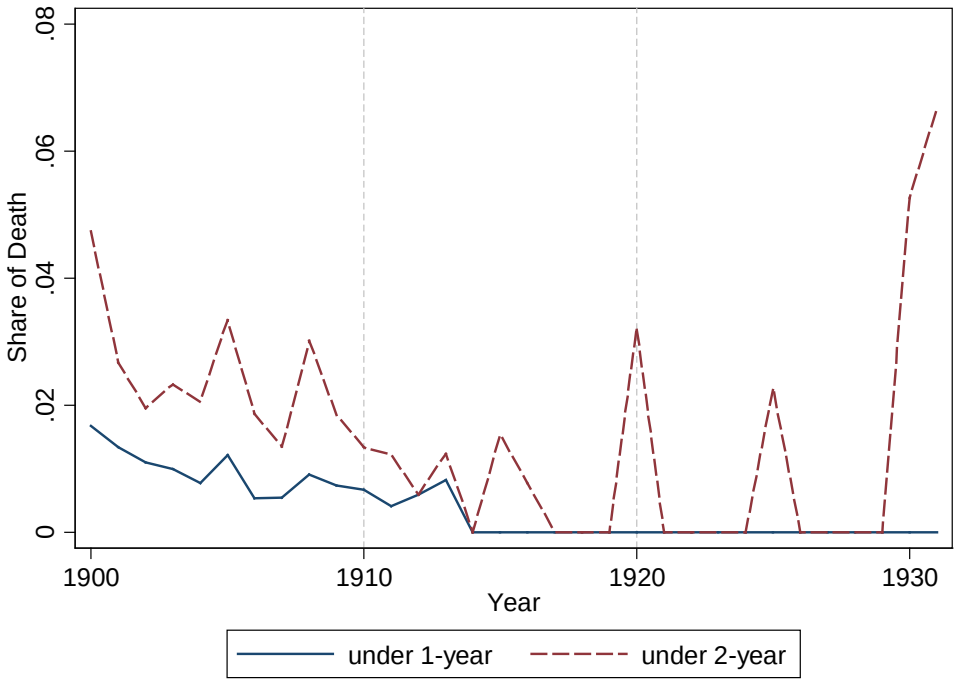


Figure 1- 7 SHARES OF DEATH IN CHILDREN UNDER ONE YEAR AND TWO YEARS OF AGE IN TOTAL TYPHOID FEVER DEATHS

In sum, diarrhea was the leading cause of death in infants and very young children in the observation period, with those under two years of age being most affected. Comparatively, diarrhea was less dangerous to children aged two years and above. From 1900 to 1915, there was an increasing trend of diarrhea-related deaths in the lower age groups (under one and two years of age). This trend was reversed in the three-year old and above age group (similar to typhoid fever), and a continuation of this declining trend was seen until 1916.

Historical facts suggest that the increasing trend of diarrhea related deaths among children under two years of age was due mainly to a decline in breast-feeding and a rise in bottle feeding involving contaminated milk. Wolf (2003) examined breastfeeding rates and early childhood mortality in Chicago from the late 1800s to the early 1900s. Wolf found that the practice of bottle feeding rose at beginning in the second half of 1800s. Traditionally, women would breastfeed their children until their second summer. After the 1880s, mothers increasingly shifted to cow's milk to feed their infants; this practice was observed in all classes. Rich and middle-class women simply desired an alternative to breastfeeding. In contrast, for economic concerns, working-class women often left infants with their older siblings, who had to offer bottled milk at feedings[31] As a result, diarrhea became an increasingly prevalent and serious health threat to infants and young children. For example, the Chicago Department of Health estimated that the death rate of bottle-fed babies was 15 times higher than that of breastfed babies (Davis 1910).[32] In the late 1800s, researchers found that the infant diarrhea-related mortality rate in Baltimore was much lower if

[31] Wolf, ibid, p. 12.
[32] Davis, "Breast Feeding," p. 2.

mothers spent more time with their children (Preston and Haines 1991). In Chicago, it was found that educating young mothers on the benefits of breastfeeding also helped reduce diarrhea-related mortality. Similar cases exist in other countries. Vögele and Woelk (2002) noticed an unexpected drop childhood diarrhea-related mortality in Berlin and other German cities during WWI, when the city milk supply was interrupted.[33]

The preceding analysis implies that i) pasteurization helped to control diarrhea epidemics, and ii) infant and early childhood diarrhea mortality does not appear to be sensitive to water-borne illness as a confounding factor. For example, I found that Chicago's diarrhea-related deaths of infants and children were largely unaffected by water quality changes in the early 1900s, but then declined after 1916 when the city adopted its full pasteurization ordinance. Below, I give some further evidence that the pasteurization intervention was a structural break in the trajectory of Chicago's diarrhea-related mortality.

Figures 8–11 illustrate that the impact of diarrhea as a cause of death in children under one year of age was much larger than in other groups. The structural change in the trajectory of diarrhea-related mortality due to the 1916 ordinance is obvious. In the figures, the diarrhea-related mortality rate is measured as the number of diarrhea-related deaths per every 100,000 children under one year of age and the typhoid fever mortality rate is the number of deaths per 100,000 of the general population.

[33] Vögele and Woelk, ibid, p. 593–594.

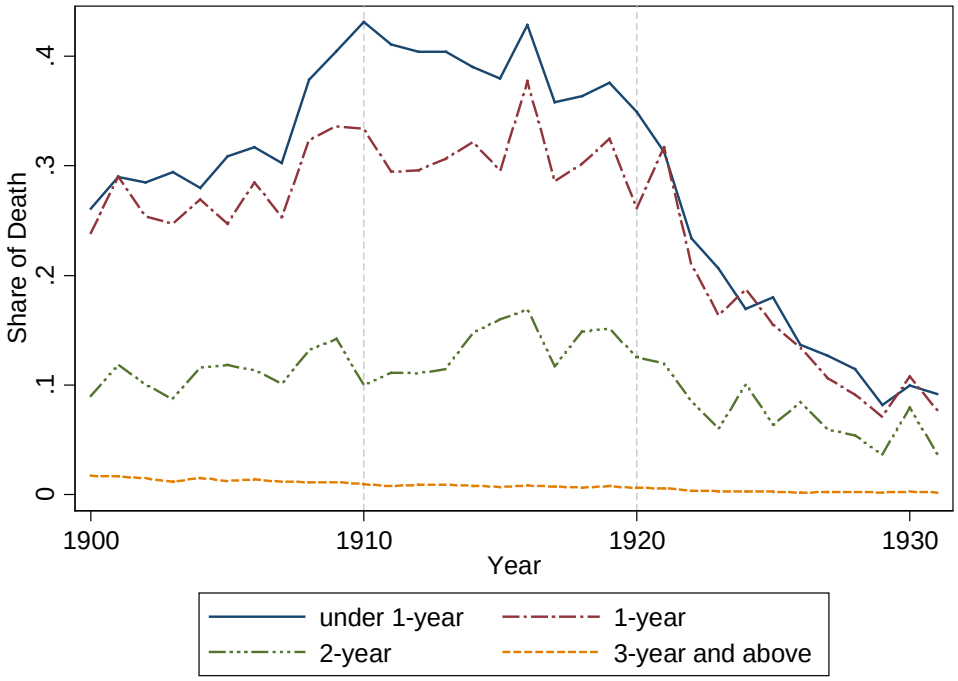


Figure 1- 8 SHARE OF DIARRHEA-RELATED DEATHS IN TOTAL MORTALITY BY AGE GROUP

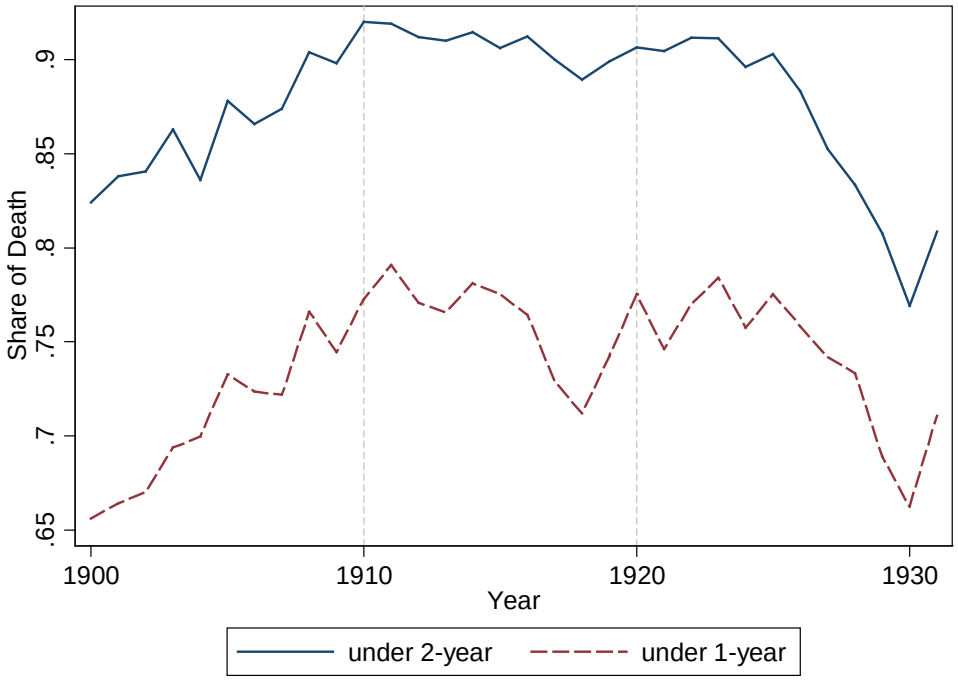


Figure 1- 9 SHARES OF TOTAL DIARRHEA-RELATED DEATHS IN CHILDREN UNDER ONE AND UNDER TWO YEARS OF AGE

Figure 1- 10 NUMBER OF DIARRHEA-RELATED DEATHS IN CHILDRER UNDER ONE-YEAR AND UNDER TWO-YEARS OF AGE

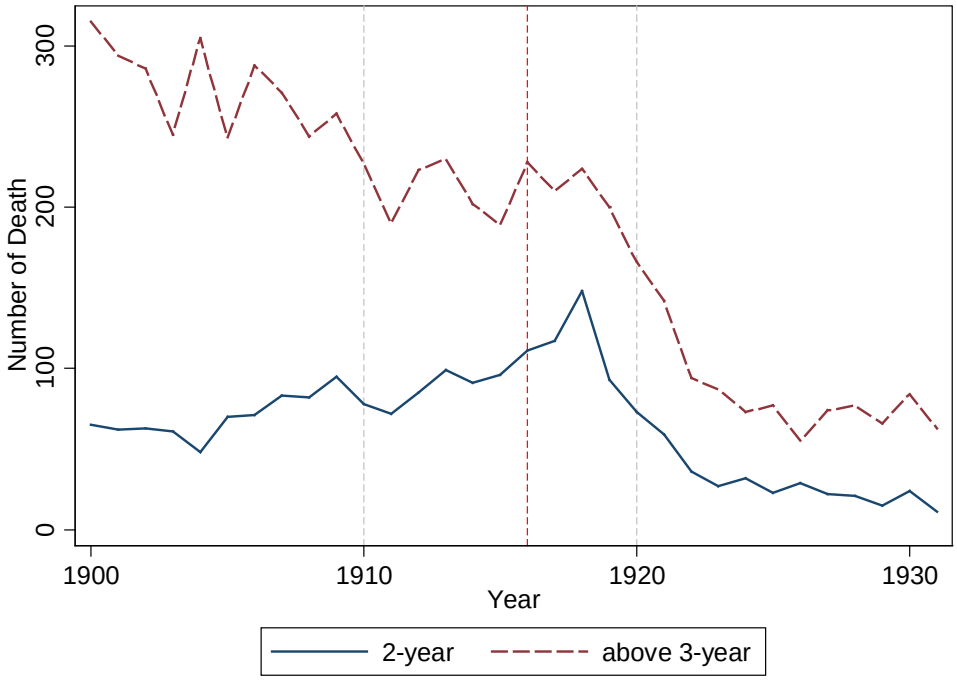


Figure 1- 11 NUMBER OF DIARRHEA-RELATED DEATHS IN TWO-YEAR OLD CHILDREN AND THOSE THREE YEARS OLD AND ABOVE

Graphically, the 1916 policy intervention occurred as a break in the series of diarrhea-related mortality rate in children under one year of age, the number of deaths in children under one year of age, and the number of deaths in children under two years of age. There was not a clear break in the number of deaths in two and three-year old children and above. These findings, plus results presented in Figures 10 and 11, explain why choosing a proper variable to reflect the health effects of our target policy are essential. Please see the Appendix for a graph that was plotted by Chicago's Health Department.[34] At that time, health officials noticed the connection between pasteurization and diarrhea, typhoid fever, and other infectious diseases. I see that the share of pasteurized milk fluctuated from 90 to 65 percent in 1914–1915, but almost all market milk was pasteurized after the 1916 ordinance was implemented.

In prior sections, I mention that both milk and water quality could affect typhoid fever-related mortality. Previous quantitative studies focused more on water purifications techniques such as filtration and chlorination as a reason for the decline in typhoid fever-related deaths. The role of milk was not included in those studies. Chicago began water chlorination 1912, largely finishing the process in 1917 (Ericson 1918).[35] This overlapped with Chicago's 1916 milk ordinance. There was a trend change of diarrhea death in 1918, which may be due to both milk and water measures, but there was less evidence that the decline in infant diarrhea-related death was correlated with water chlorination (1918 was also the year of the Spanish Influenza epidemic). In the 1900-1910, the trend of infantile diarrhea death was opposite that of typhoid fever mortality change.

[34] Robertson, ibid, p. 1002.
[35] Ericson, "Chlorination of Chicago's Water Supply," p. 251.

4. CAUSAL EFFECTS ANALYSIS

Regression-Based Methods

The above analysis identifies what seems to be a causal relation between diarrhea-related mortality and milk pasteurization in Chicago and confirms that the drop in typhoid fever-related mortality was largely due to water purification measures. This section attempts to measure the causal health effects of the 1916 policy change.

One major empirical challenge is the choice of a proper empirical method. The causal effect of (D) is defined as $C(D, \mathbf{X}, e) = \nabla f(D, \mathbf{X}, e)$, [36] holding vector $\mathbf{X}$ and unobserved component e constant. For Chicago (i), the impact of the 1916 ordinance is measured as (1).

$$\nabla f(D, e, X) = (y_i \mid \mathbf{X}_i, e_i, D_i = 1) - (y_i \mid \mathbf{X}_i, e_i, D_i = 0 \qquad (1)$$

That is the Average Treatment Effect (ATE). However in economic studies, the validity of ATE is threatened by complications such as correlation between outcomes and treatment, omitted variables bias, and endogeneity of treatment variables. In practice, counterfactuals are usually applied to make statistically meaningful estimations (Cameron and Trevidi 2005).[37] In this study, there are two options:

(1) If there is no concurrent trend, I can use the before-and-after design to measure the treatment effect, and no control group is needed. This is the approach that Cutler and Miller (2005) used in their study of the health effects of water treatment. The authors assumed potential cofounding changes are common across treated units, and city-specific conditions remained the same across the period. The only variation came from changes in water treatments. Here y_{i0} and y_{i1} are the outcomes when $D = 0$ and

[36] Hansen, *Econometrics*, p. 43.
[37] Cameron and Trevidi, *Microeconometrics: Methods and Applications*, p. 32–33.

$D = 1$; γ_i is the individual fixed effect; λ_t is the time effect covering two periods, namely the pre-treatment period with $t = 0, \forall D = 0$ and the post-treatment period with $t = 1, \forall D = 1$; and $m_t(\mathbf{X}_i)$ is the function of other control variables. In the format of the Potential Outcome Model (POM), the treatment effect can be measured as follows:

$$\begin{cases} y_{i0} = \gamma_i + \lambda_0 + m_0(\mathbf{X}_i) + e_i & \forall D = 0 \\ y_{i1} = \gamma_i + \lambda_1 + m_1(\mathbf{X}_i) + e_i & \forall D = 1 \end{cases} \quad (2)$$

If $m_1(\mathbf{X_i}) = m_0(\mathbf{X_i}) + \alpha$, then $y_{i1} = \gamma_i + \lambda_1 + m_0(\mathbf{X_i}) + \alpha + e_i$.

The treatment effect is:

$$\alpha = (y_{i1} - y_{i0}) - (\lambda_1 - \lambda_0) \quad (3)$$

In the case of no concurrent trend or $\lambda_1 - \lambda_0 = 0$, then $\alpha = y_{i1} - y_{i0}$.

(2) If the assumption $\lambda_1 - \lambda_0 = 0$ is not satisfied, an alternative is to choose a group of control units (*j*) with a similar time trend as Chicago.

$$\begin{cases} y_{j0} = \gamma_j + \lambda_0' + m_0(\mathbf{X}_j) + e_j & \forall D = 0 \\ y_{j1} = \gamma_j + \lambda_1' + m_1(\mathbf{X}_j) + e_j & \forall D = 1 \end{cases} \quad (4)$$

$y_{j1} = \gamma_j + \lambda'_1 + m_0(\mathbf{X_j}) + e_j$, as no treatment occurred in the control group.

Thus, $y_{j1} - y_{j0} = \lambda'_1 - \lambda'_0$, if $m_1(\mathbf{X_j}) - m_0(\mathbf{X_j}) = 0$.

If the common trend assumption is satisfied $\lambda'_1 - \lambda'_0 = \lambda_1 - \lambda_0$, then the treatment effect will be Equation (5).

$$\alpha = (y_{i1} - y_{i0}) - (y_{j1} - y_{j0}) \quad (5)$$

The previous literature has identified the significance of the common trend assumption in empirical DD models. Wolfers (2006) indicated inconsistency of the

treated and control units before intervention could alter policy evaluation results. In her study of retail gasoline prices and competition, Hastings (2004) used parallel pre-treatment price trajectories to ensure the common trend assumption is satisfied. In this study, the DD model also requires control units to have had a similar dependent variable trend as Chicago before 1916.[38]

Data

In the POM framework, a comparison group with proper control units is the key to estimating the causal treatment effect. In this study, the control unit for Chicago should be a city without a milk pasteurization ordinance during the study period. The control group used here is comprised of the 20 US cities with populations above 100,000 (in 1930) that had no mandatory pasteurization ordinance by 1931, as identified in a survey by Frank and Moss (1931). None of the cities had more than 90 percent of its milk pasteurized by 1931, while Chicago achieved 99 percent pasteurization by 1924.[39]

Next, I consider an appropriate dependent variable to reflect the effect of this policy intervention. In the prior section, the time series discussion is based on the diarrhea-related mortality rate and the number of diarrhea-related deaths in children under one-year old. In this part, the second variable (number of diarrhea-related deaths) is less appropriate, since population varied significantly across cities. Thus, a comparable variable is mortality rate. However, the total number of diarrhea-related deaths among children under one-year old and the population share of this age group

[38] Wolfers, "Did Unilateral Divorce Laws Raise Divorce Rates," p. 1802–1820; Hastings, "Vertical Relationships and Competition in Retail Gasoline Markets," p. 317–328.
[39] Frank and Moss, "The Extent of Pasteurization and Tuberculin Testing in American Cities," p. 1–4.

were not available in some small cities. For many control units, the mortality statistics contained only data on diarrhea-related deaths in children under two years of age. However, children under the age of two were not recorded in decennial census. As an alternative, I calculate the mortality rate as a ratio of the number of diarrhea-related deaths of children under two years of age to every 1,000 people under five years old.[40] Annual population data were not available. In this study, they were averaged using 1900, 1910, 1920, and 1930 census data. To control for possible influences on death rates, I included demographic covariates and income, including female share, white share, and average wage in manufacturing, which were obtained from decennial 1900 to 1930 censuses.

Single Unit Pre/Post-Treatment Comparison: Chicago and St. Paul

Figure 12 shows the pre-treatment outcomes of Chicago and 20 control cities. It is clear that the trajectory of Chicago (solid line) is different from most of the control units (dash lines) in the pre-treatment period. Also, the arithmetic average of the control group was also not similar to Chicago before 1916 (Figure 13). In the control group, only St. Paul's trend approximated Chicago's pattern (Figure 14). They were largely parallel from 1910–1915 (Figure 15).

[40] The share of population under five years of age was available in all cities in 1910. But it was not available in some cities in 1900. For these missing values, it was imputed by the correlation between 1900 and 1910. The coefficient is 0.986, which implies the share of children under five years of age remained stable from 1900 to 1910.

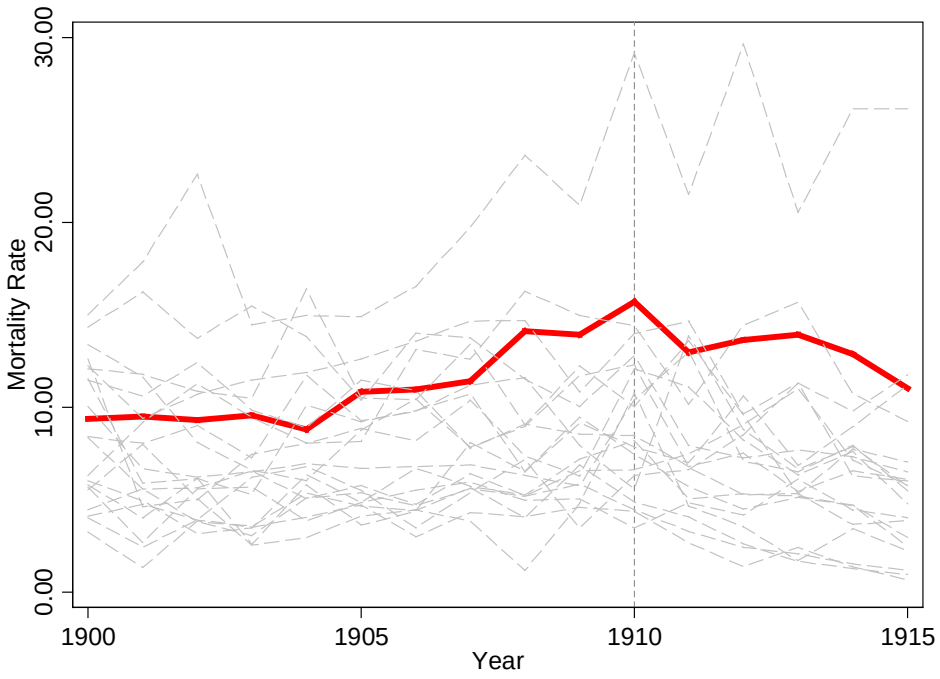


Figure 1- 12 CHILDREN DIARRHEA MORTALITY RATES OF CONTROL UNITS (DASH) AND CHICAGO (SOLID), 1900–1915

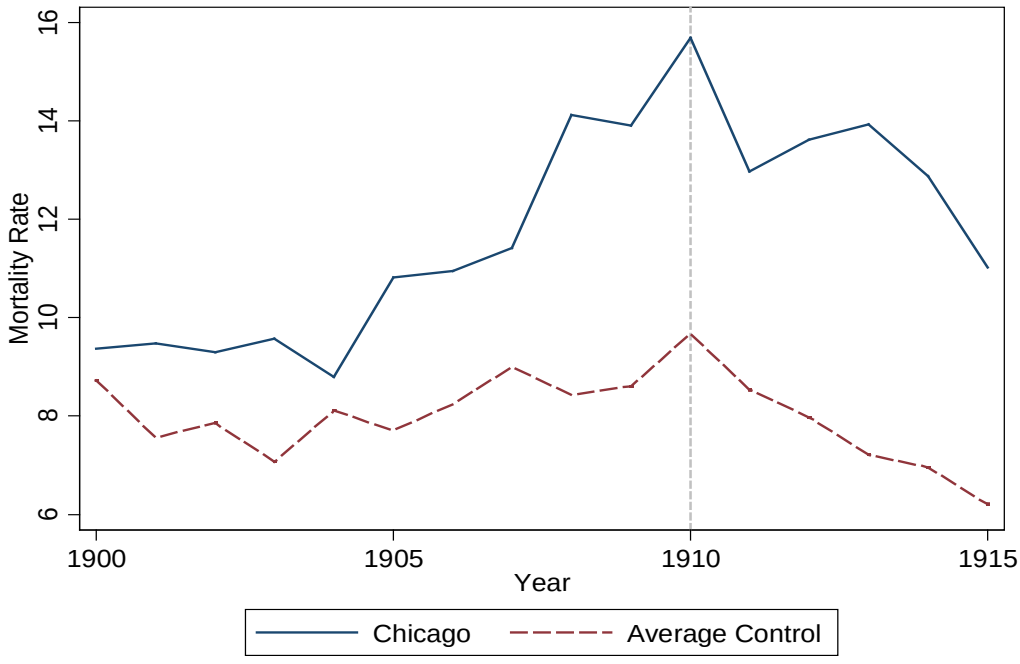


Figure 1- 13 MORTALITY RATES OF THE AVERAGED CONTROL GROUPS AND CHICAGO, 1900–1915

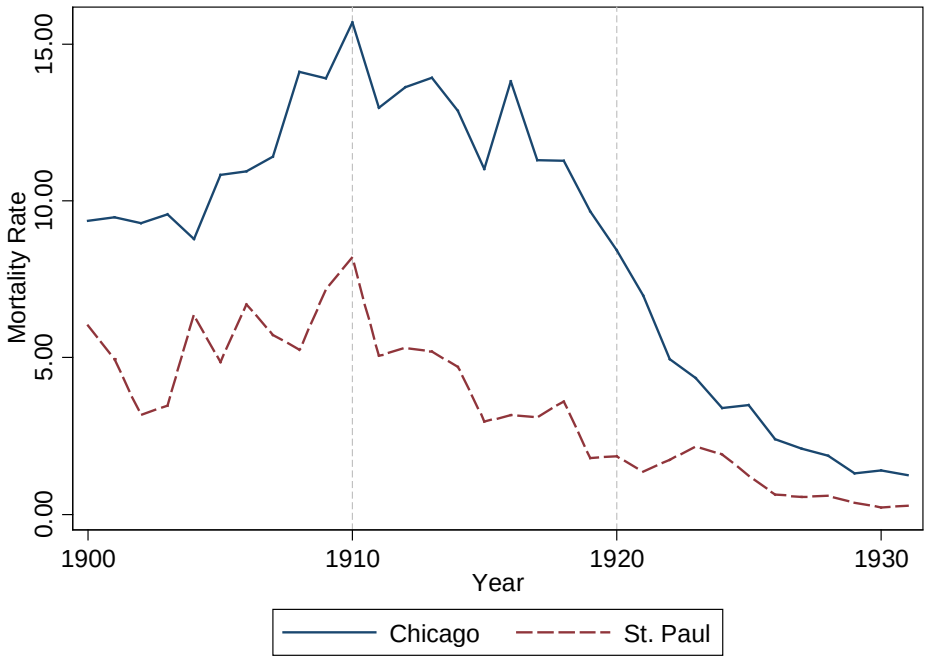


Figure 1- 14 MORTALITY RATES OF CHICAGO AND ST. PAUL 1900–1931

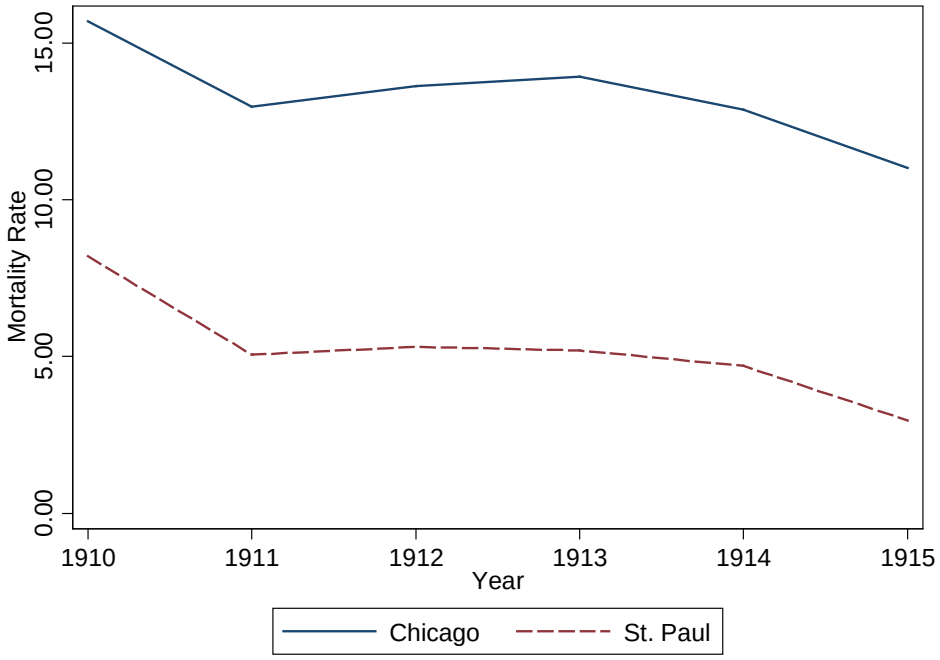


Figure 1- 15 MORTALITY RATES OF CHICAGO AND ST. PAUL 1910–1915

Using St. Paul as a counterfactual, I made a DD comparison to the causal effect of the 1916 intervention. The gaps in Chicago and St. Paul's outcomes are summarized in Table 1, which illustrates their difference was roughly 8.0 from 1910–1915. Holding other factors constant, the difference between Chicago and St. Paul after 1916, minus the 1910–1915 gap, can be used to demonstrate the causal effect of the 1916 ordinance. Bertrand et al. (2004) pointed out that many economic outcomes are correlated over time, and the error components can be serially correlated. To remedy this, I cluster the outcomes into two parts, the average of the pre-treatment period (1910–1915) and the average of the post-treatment period (1916–1931). Figure 16 illustrates the design of this DD comparison. The estimated effect is $\alpha = (\bar{y}_{i1} - \bar{y}_{i0}) - (\bar{y}_{j1} - \bar{y}_{j0})$ = 4.15.

This DD model with a single comparison unit highlights the positive health effects of the 1916 ordinance. However, some DD models may have two weaknesses.[41] First, there may be uncertainty about values of aggregate variables in the population. Second, there may be ambiguity of how the comparison unit is chosen. In this study, the first uncertainty is not a concern, since I use aggregated city-level data instead of a sample of disaggregated units. But the second ambiguity cannot be ignored. In the DD model, St. Paul was chosen due to its similarity to Chicago in 1910–1915. However, additional comparison cities should be used if I observe other "quantifiable characteristics." In the next section, I use a non-parametric data driven process to identify a better control group.

[41] Abadie, Diamond, and Hainmueller, "Synthetic Control Method for Comparative Case Study: Estimating the Effect of California's Tobacco Control Program," p. 493–494.

Table 1- 1 PRE- AND POST-TREATMENT OUTCOME DIFFERENCES OF CHICAGO AND ST. PAUL 1910–31

Year	Chicago	St. Paul	Diff.	Year	Chicago	St. Paul	Diff.
1910	15.69	8.20	*7.49*	1921	6.99	1.37	5.62
1911	12.97	5.06	*7.91*	1922	4.95	1.74	3.21
1912	13.63	5.31	*8.32*	1923	4.35	2.17	2.18
1913	13.93	5.20	*8.73*	1924	3.39	1.91	1.48
1914	12.88	4.70	*8.18*	1925	3.48	1.23	2.25
1915	11.02	2.97	*8.05*	1926	2.39	0.64	1.75
1916	13.83	3.16	10.67	1927	2.10	0.55	1.55
1917	11.30	3.10	8.20	1928	1.87	0.59	1.28
1918	11.29	3.61	7.68	1929	1.31	0.37	0.94
1919	9.67	1.80	7.87	1930	1.40	0.23	1.17
1920	8.42	1.86	6.56	1931	1.24	0.28	0.96

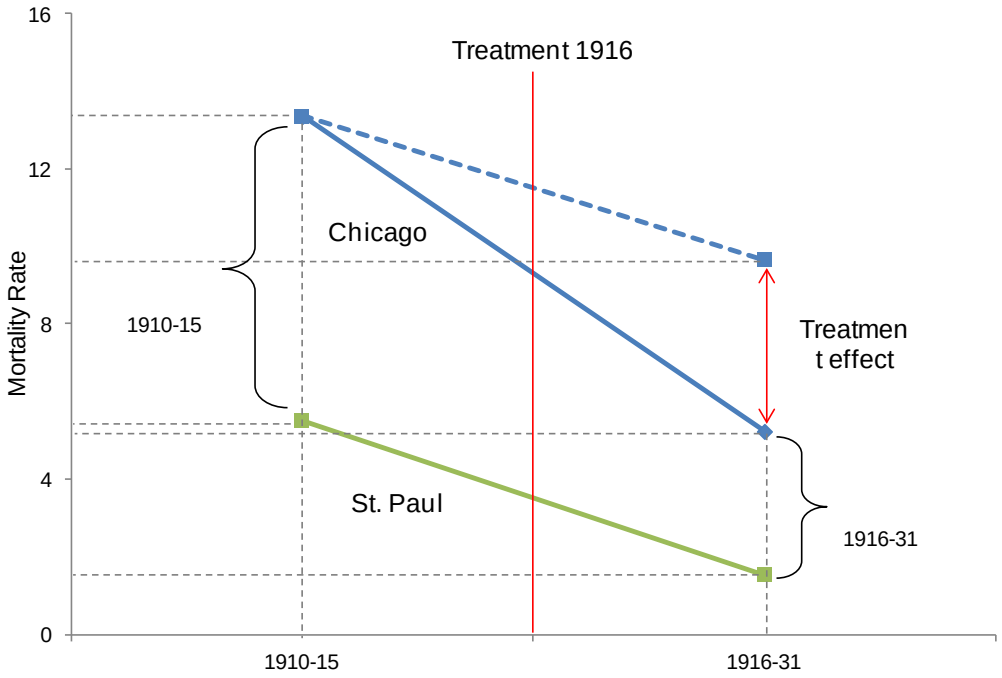


Figure 1- 16 DIFFERENCE-IN-DIFFERENCE SET UP OF CHICAGO AND ST. PAUL 1910–1931

5. A SYNTHETIC CONTROL STUDY

Model Set-Up

The synthetic control method (SCM) was developed by Abadie and Gardeazabal (2003), and its specifications and algorithms were more formally derived in an application to California's tobacco cessation program by Abadie *et al*. (2010). The insight behind SCM is that matching a single treated unit in a case study with a convex combination of comparison units yields a synthetic or counterfactual version of the treated unit. The treatment effect can then be estimated by differencing the outcome for the treated unit in the post treatment period against the outcome for its synthetic self, which was not exposed to policy intervention. Statistical significance is estimated by constructing synthetic counterfactual units for all members of the control group in order to identify the distribution of the estimator under the null hypothesis (of no effect).

The synthetic version of the treated unit is a convex combination of control units optimized by minimizing the distance between the real unit and its synthetic version in the pre-intervention period. Synthetic versions of control units are generated similarly, but disallow any weight from the treated unit itself. Distance is measured as the Euclidean distance between vectors comprised of covariates and pre-treatment outcomes. By allowing matching on pre-treatment outcomes in addition to covariates, SCM is robust to violations of the constant trend assumption, which DD and FE estimators are unable to handle.

The SCM model is specified as follows. Suppose there is one treated unit, I, and N control units, j ($j = 1, 2 \ldots N$). I consider a policy intervention with data

sampled both before and after treatment. The pre-treatment periods are $t = t_0, \dots, t_k$, and the post-treatment periods are $t = t_{k+1}, \dots, T$, so treatment happens between periods t_k and t_{k+1}. Let Y_{it} denote an outcome in t for the treated unit, and let Y_{jt} denote an outcome in period t for control unit j. **X** is a vector of predictors (covariates). For i the treatment effect, α_{it}, is measured as the difference between its post-treatment outcome, Y_{it}, and its synthetic post-treatment outcome, Y^*_{it}, which is given by a convex combination of the post-treatment outcomes of control units, Y_{jt}, defined by optimized weights, w^*_j,:

$$y_{it} = \beta_i \cdot X_{it} + \alpha_{it} \cdot D_{it} + e_{it} \quad \text{where } D_{it} = \begin{cases} 0, & t = t_0, \dots, t_k \\ 1, & t = t_{k+1}, \dots, T \end{cases} \tag{6}$$

$$\begin{aligned} & y_{jt} = \beta_j \cdot X_{jt} + e_{jt} \quad \text{and} \\ & y_{it}^{*} = \sum_{j=1}^{N} w_j^* \cdot y_{jt} = \sum_{j=1}^{N} w_j^* \cdot [\beta_j \cdot X_{jt} + e_{jt}] \end{aligned} \tag{7}$$

The treatment in the SCM model is the difference between the real treated unit and its synthetic version after the treatment as:

$$\begin{aligned} & \alpha_{it} = Y_{it} - Y_{it}^{*} = Y_{it} - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \quad \forall t \geq t_{k+1} \text{ and} \\ & \text{s.t. } w_j^* \geq 0 \text{ and } \sum_{j=1}^{N} w_j^* = 1 \end{aligned} \tag{8}$$

The optimized weights, w_j^*, are obtained by minimizing the distance M between $\mathbf{X}_j$, and $\mathbf{X}_j \cdot \mathbf{W}_j$ in the pre-intervention periods, according to:

$$M = \min_{w_j} [(\mathbf{X}_i - \mathbf{X}_j \cdot \mathbf{W}_j)' \mathrm{V} (\mathbf{X}_i - \mathbf{X}_j \cdot \mathbf{W}_j)]^{1/2} \quad \forall t \in (t_0, t_k) \tag{9}$$

where the matrix, ***V***, is positive definite and chosen to minimize the mean squared prediction error (MSPE) with respect to pre-treatment outcomes only, conditional on values of w^*_j. To be clear, this process is what distinguishes SCM from a DD

approach, because control units are weighted according to the optimized w_j^*, instead of a simple weighting of $w_j = 1/N$. Recall Figure 13, which shows that the averaged control group is not a good counterfactual to Chicago, since their trajectories were not following a similar trend before the treatment. Obviously, their pre-treatment trends were not close. That is one rationale to use SCM. In addition, causal effects are obtained holding other factors constant. In a regular regression framework, these factors are controlled as covariates on the right hand side (RHS). This point is challenging in our study, since all demographic and income factors are from decennial census years, but our dependent variable is yearly. Thus, the second reason to apply SCM is that it can transform the influence of decennial control variables into the optimized weights of the comparison units.

Estimations and Results

The SCM estimation algorithm includes the following steps:

1. Construct a synthetic version of Chicago using 20 control cities and evaluate the gap between treated Chicago vs. synthetic, untreated Chicago. The real-synthetic gap of Chicago estimates the actual health effects of the 1916 ordinance.

2. Construct a synthetic version of each of the 20 control cities using the remaining 19 control cities and evaluate each city against its synthetic version. Treatment was also imposed in 1916. As there was no treatment, the 20 counterfactual gaps measure the hypothetical gaps under the null hypothesis of no health effect.

3. Calculate the root of mean squared prediction errors (RMSPE) in the pre-treatment period (RMSPE1) and post-treatment period (RMSPE2). RMSPE1 is an indicator of

pre-treatment fitting, and RMSPE2 measures treatment effects. The ratio RMSPE2/ RMSPE1 reflects the treatment effects.

The study of Dube and Zipper (DZ 2013) [42] indicated that SCM users should be careful when choosing covariates (predictors) to ensure matching quality. The optimized weights of control units and synthetic versions are determined by predictors,[43] but explicit predictor selection guidance is not always available. DZ proposed a five-step approach for an optimal set of predictors via cross-validation. First, for each set of predictors, the pre-intervention observations are used to select optimal donor weights and predictor weights. Second, those weights and the post-intervention observations are used to calculate each predictor set's prediction error for each of the *N-1* donors, where N is the total number of control units. Third, the sum of squared post-intervention predictor error is calculated for each control units. Fourth, for each predictor set, the post-intervention prediction errors are averaged by *N-1*. In the last step, researchers choose the prediction set that has the lowest sum of squared errors.

$$RMSPE = \left\{ \frac{1}{t_1} \sum_{t_1=1}^{t_1=t_k} [Y_{it}(1) - Y_{it}(0)]^2 \right\}^{1/2} \quad (10)$$

In this study, without a formal predictor selection algorithm as DZ above, I use all pre-intervention outcomes as predictors for simplicity. This approach may not lead to the best set of predictor. However, as DZ explained, it will lead to the lowest root of mean squared prediction errors (RMSPE) as in Equation 10. At the same time, other covariates would become redundant when all pre-intervention outcomes are included.

[42] Dube and Zipper, "Pooled Synthetic Control Estimates for Recurring Treatments," p. 12–14.
[43] Also, a different set of control units leads to different synthetic versions with the same predictors.

Table 2 reports the optimized weights which are generated from using all pre-intervention outcomes as predictors.

Table 1- 2 RMSPE AND CONTROL UNIT WEIGHTS OF SCM SPECIFICATIONS

City	Weight	City	Weight	City	Weight
Duluth	0.077	Nashville	0	San Diego	0.148
Evansville	0.009	New Haven	0	Seattle	0
Hartford	0	New Orleans	0.068	St. Paul	0
Jacksonville	0	Omaha	0	Tacoma	0
Kansas city	0.004	Portland	0.041	Utica	0.302
Los Angeles	0	Providence	0	Wichita	0
Memphis	0.125	San Antonio	0.227		

Graphically, Figure 17 illustrates the SCM result (Model 1). The solid line represents the real mortality rate of Chicago. The synthetic Chicago is indicated by the dashed line. The vertical dashed line marks the 1916 policy intervention. Recall if the policy of 1916 was effective, the mortality rate in Chicago should be lower than its synthetic version (no treatment). In other words, there should be a negative real-synthetic gap. Given a good fit between real and synthetic Chicago before 1916, the larger negative real-synthetic gap is the greater health impact of the 1916 ordinance. As shown in Figure 17, there was a noticeable negative real-synthetic gap after 1920, which means Chicago had better health outcomes. Overall, after 1919 the decline in Chicago's mortality rate was steeper than that of the optimized control group. For example, Chicago's mortality rate dropped from 9.67 in 1919 to 6.99 in 1921, about 27.8% drop, while mortality rate in the linear weighted control units increased 19.5%, from 6.58 to 7.87. The trend continued in the early 1920s. Mortality rate in Chicago

was 3.48 in 1925 and 50.2% lower than 1921. For the counterfactual, its decline was only 11.1% and dropped from 7.87 to 6.99.

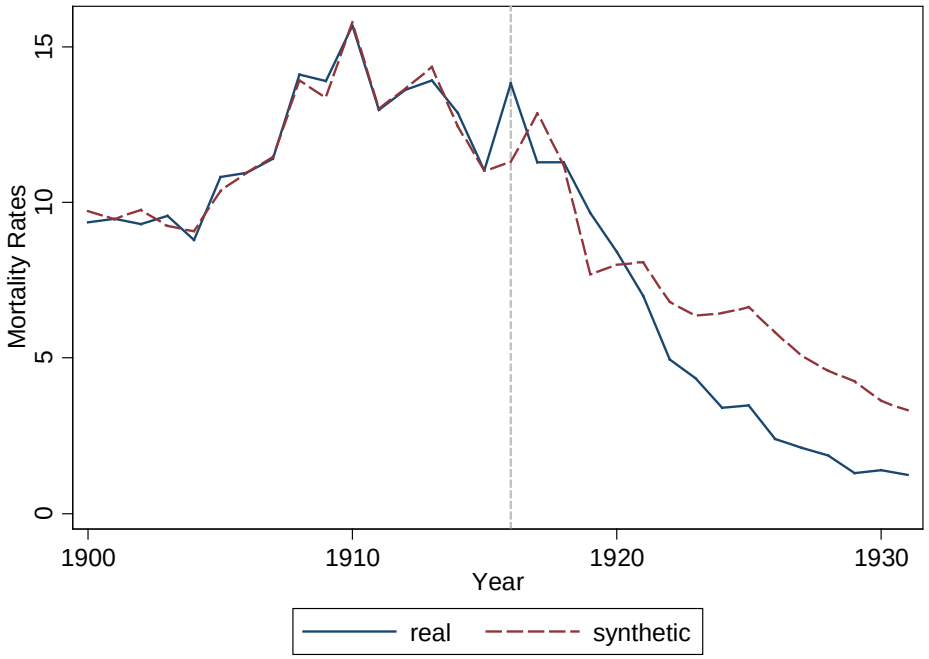


Figure 1- 17 CHICAGO—SYNTHETIC VERSION (MODEL 1)

Inferences and Robustness Tests

I obtained a substantial negative real-synthetic gap driven by the SCM algorithm as above. However, the gap, by itself, cannot guarantee statistically significant health effects of the ordinance. There is possibility that this effect was driven by pure random chance.[44] In other words, this gap could be even bigger between real and synthetic control versions of an unexposed control unit. So inference tests are needed to prove whether the effect was meaningful at normal statistical levels. Following Abadie *et al*. (2010), I chose "placebo studies" and make inferences, which randomly reassign the intervention to all control cities. If the effect was not from purely random chance, the effect should be noticeably different from its comparison units.

[44] Adabie, Diamond, and Hainemuller, ibid, p. 501.

To implement it, I conduct a series of placebo studies using the iterative SCM process. For each control unit, its synthetic control version is constructed using the remaining 19 control cities in the donor pool. The predictors are the same as the SCM of Chicago. The 20 real-synthetic gaps generated from the above iterative process are then considered to be the nonparametric distribution of the gaps under the null hypothesis. Figure 18 illustrates that Chicago was lower than most control units from the middle 1920s. Before that, Table 3 shows treatment effect of Chicago did not stand out in the comparison group. But the effects became noticeable after 1925. For example, in 1925-1929, Chicago's gap was larger than all 20 control units. In other words, the probability that a control city could surpass Chicago was only $1/21 \approx 0.048$, which is akin to the *p*-value in a conventional statistical summary report.[45]

As Abadie *et al*. (2010) pointed out, another concern is post-treatment gaps, which may be generated from the lack of fitting between pre-treatment real and synthetic trajectories.[46] To verify this inference, I need to consider the real-synthetic fitting before and after the intervention. Here I chose the post/pre-treatment RMSPE ratio to reflect the effect of the 1916 intervention. The ratio is calculated as:

$$\sigma^{MSPE} = \left[\frac{1}{t_2}\sum_{t_2=t_{k+1}}^{t_2=T}[Y_{it_2}(1) - Y_{it_2}(0)]^2\right] \Big/ \left[\frac{1}{t_1}\sum_{t_1=1}^{t_1=t_k}[Y_{it_1}(1) - Y_{it_1}(0)]^2\right] \qquad (11)$$

According to the principle of SCM, a small pre-treatment MSPE ($t = t_0, \ldots, t_k$) is good, indicating a good fit between real and synthetic trends. A large post-treatment MSPE ($t = t_{k+1}, \ldots, T$) indicates noticeable intervention effects. Figure 19 shows that the ratio of Chicago (3.64) is higher than all control units, which is an illustration of

[45] In this case, one control unit had a real-synthetic gap lower than Chicago; the probability was $2/21 \approx 0.095$. Considering the small sample size, the 10% significance level is still acceptable.
[46] Abadie, Diamond, and Hainemuller, 2010, ibid, p. 502.

the noticeable health effects of the 1916 ordinance. The probability of the significance level obtaining a post/pre-treatment RMSPE ratio lower than Chicago is $1/21 \approx 0.048$, as above. The health effect of Chicago's ordinance is statistically significant.

Table 1- 3 THE NUMBER OF CONTROL UNITS THAT HAD REAL-SYNTHETIC GAPS (NEGATIVE) THAN CHICAGO

Year	Number	Year	Number	Year	Number	Year	Number
1916	NA	1920	NA	1924	2	1928	0
1917	9	1921	8	1925	0	1929	0
1918	NA	1922	2	1926	0	1930	1
1919	NA	1923	4	1927	0	1931	1

Note: NA Chicago's real-synthetic gap was positive in that year

For robustness, I replicate the process with a different control group. For example, I dropped some control units with extreme values of their dependent variable. The SCM result is similarly. The real-synthetic gap of Chicago is lower than most control units in the 1920s.[47] The result is robust.

[47] There are three possibilities that made treatment effects in Chicago noticeable only after 1920. First, as in many other large cities, the 1918 influenza epidemics raised overall mortality rates in Chicago and could offset the health effects of pasteurization. So it could slow the drop of children diarrhea deaths. Second, at absolute levels, Chicago's children mortality rates were much higher than many control units. It was not rare that children's overall mortality rates were lower in smaller cities. So even with a higher declining rate, Chicago's real-synthetic gaps may be lower than some control units. Third, as Abadie (online) suggested, treatment effect may take a while to be noticeable in SCM applications. In that case, Abadie suggest including enough post-intervention observations.

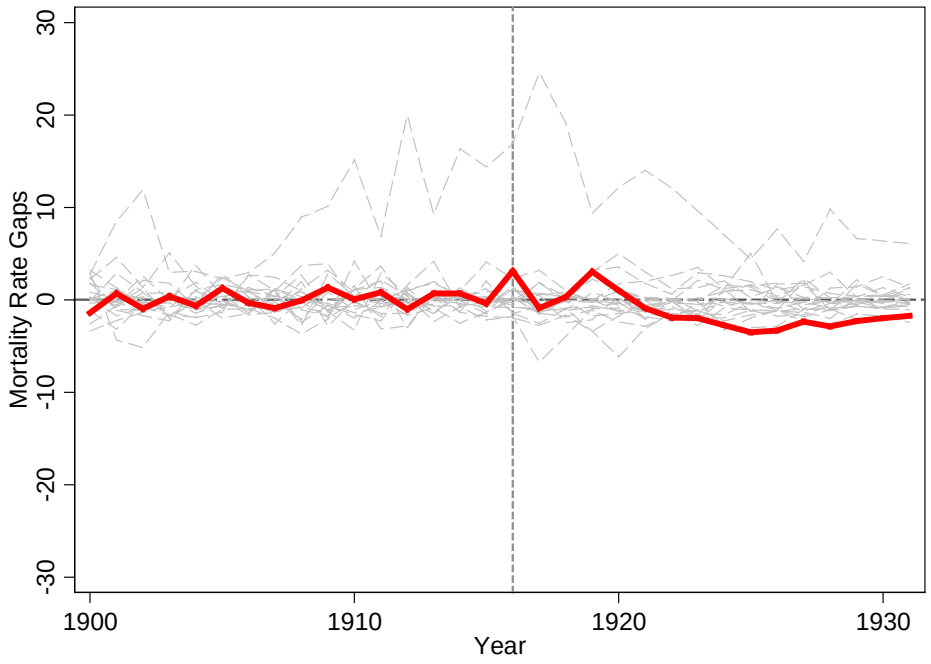


Figure 1- 18 REAL-SYNTHETIC GAPS OF CHICAGO (SOLID) AND CONTROL CITIES (DASH)

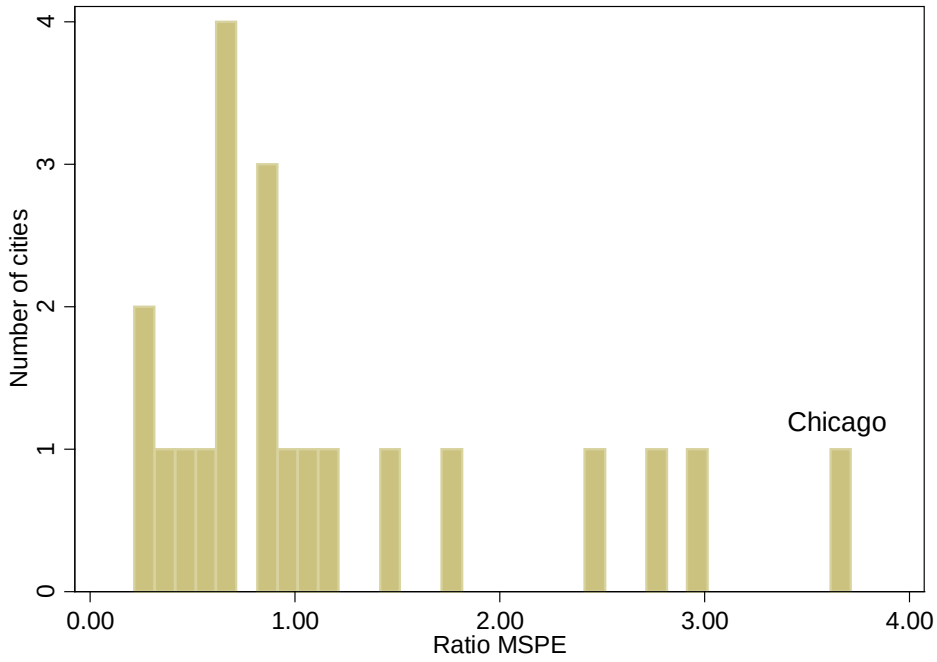


Figure 1- 19 PRE/POST MSPE RATIO OF CHICAGO AND 20 CONTROL CITIES

6. CONCLUSIONS

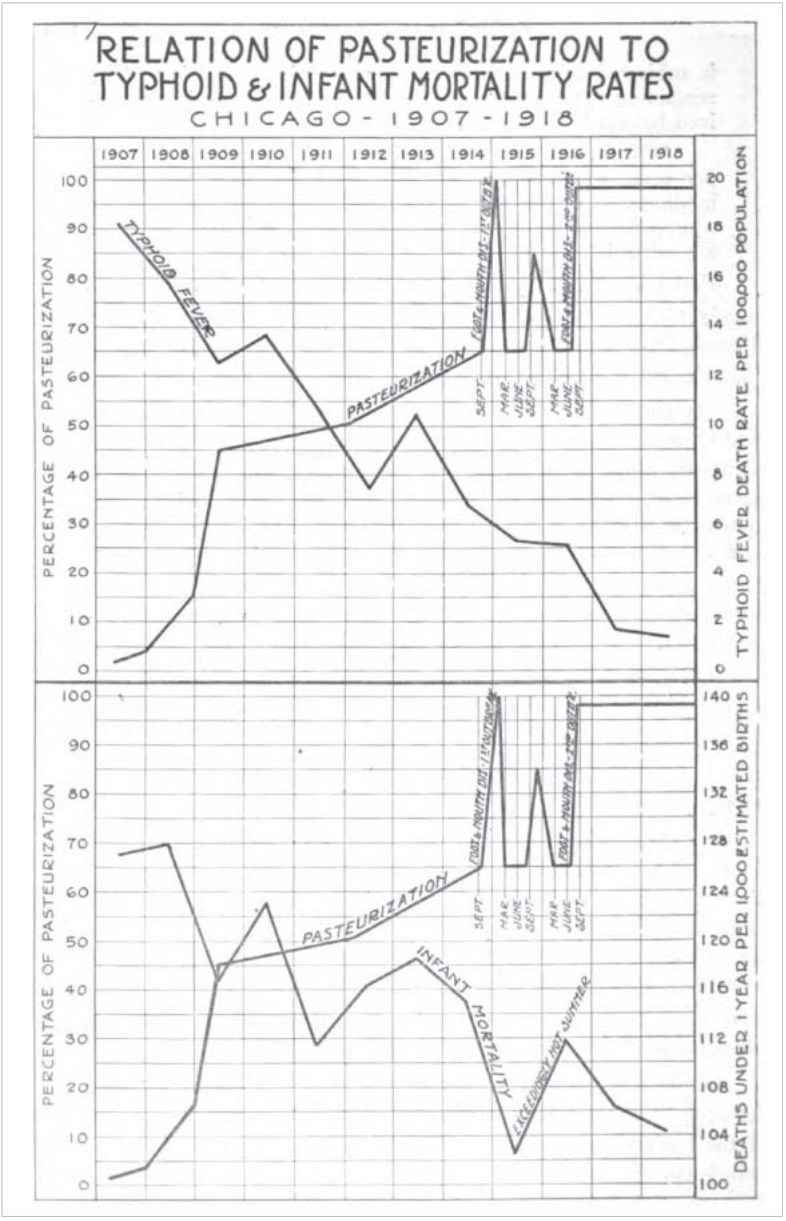
Using Chicago's 1916 pasteurization ordinance, this paper provides a case study of measuring the health impacts of food safety interventions. Empirically, there are two major challenges in estimating the effects of this policy intervention. The first is confirming the causal relationship between pasteurization and health consequences. In other words, I needed to determine if it was milk pasteurization or other factors that changed the trajectory of the dependent variable. The second is employing the proper model to capture the causal effect of the target policy intervention. In this study, the empirical strategy is constrained by data availability and the need for a proper counterfactual.

To address the first challenge, I analyzed time variations in the outcome variables of interest. The results shed light on mortality transitions over time of diarrhea and typhoid fever. They indicate that pasteurization was the leading factor in the decline of childhood diarrhea-related mortality. Water treatment was responsible for a lower mortality rate from typhoid fever, but had no direct impact on Chicago's infant and early childhood diarrhea-related deaths. Indeed, the trend of infant diarrhea-related mortality was the opposite of typhoid fever-related mortality from 1900 to 1910. In that period, the typhoid fever mortality rate decreased as a consequence of water filtration. In contrast, the diarrhea-related mortality rate of infants continued rising, since more mothers discontinued breastfeeding and shifted to bottled milk. Thus, the results suggest that typhoid fever was not a confounding factor in infant and early childhood diarrhea-related mortality. The lower childhood diarrhea mortality came from better milk quality in Chicago.

To address the second challenge, the non-parametric data driven SCM approach was applied to capture the causal health effects of this ordinance. Estimation and causal inference for this type of historical policy evaluation is challenging due to the presence of unobserved heterogeneity in the data and the problem of using appropriate comparison units (the control group). Control units with characteristics similar to Chicago before 1916 are rare (only St. Paul in 1910–1915). In addition, I also needed a strategy to use non-yearly covariates data. Following Abadie *et al*. (2010), I used the SCM process and built a synthetic Chicago with 20 control units and a set of predictors. After choosing the best predictors to minimize the distance of the two trajectories before 1916, a noticeable real-synthetic gap was observed in the post-treatment period. In 1921 to 1931, on average, Chicago's ral mortality rate was 2.31 lower than its counterfactual. In addition, the post/pre-treatment MSPE ratios suggest that the effect of this ordinance was more noticeable in Chicago than in the 20 control cities. In sum, I find Chicago's pasteurization ordinance had statistically significant, positive health effects.

APPENDIX

Relationship of Pasteurization and Typhoid Fever to Infant Mortality



Source: Roberston (1919).

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. “Synthetic Control Method for Comparative Case Study: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105, no. 490 (2010): 493–505.

------. “Comparative Politics and the Synthetic Control Method.” *American Journal of Political Science* (Article first published online: 23 APR 2014).

Abadie, Alberto, and Javier Gardeazabal. “The Economic Costs of Conflict: A Case Study of the Basque Country.” *American Economic Review* 93, no.1 (2003): 112–32.

Abadie, A., “Using Synthetic Controls to Evaluate an International Strategic Positioning Program in Uruguay” [Online], Available at: http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=37312864

Alvord H. E., and R. A. Pearson. “The Milk Supply of Two Hundred Cities and Towns.” U.S. Department of Agriculture, Bureau of Animal Industry, Bulletin, 46, Washington, DC: 1903.

Atkins Peter J. The Pasteurization of England: The Science, Culture and Health Implications of Milk Processing, 1900–1950. In *Food, Science, Policy and Regulation in the Twentieth Century: International and Comparative Perspectives*, edited by David F. Smith and Jim Phillips, London: Routledge, 2000: 37-51

Atkins, Peter J. *Mother’s milk and infant death in Britain, circa 1900–1940. Anthropology of Food* [Online]. Available at: http://aof.revues.org/index310.html. 2003.

Beaver, M. W. “Population, Infant Mortality, and Milk.” *Population Studies* 27, no. 2 (1973): 243–54.

Block, Daniel R. “Purity, Economy, and Social Welfare in the Progressive Era Milk Movement.” *Journal for the Study of Food and Society* 3, no.1 (1999): 20–27.

------. “Public Health, Cooperatives, Local Regulation, and the Development of Modern Milk Policy: The Chicago Milkshed, 1900–1940.” *Journal of Historical Geography* 35, no. 1 (2009): 128–53.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. “How Much Should We Trust Differences-In-Differences Estimates?.” *The Quarterly Journal of Economics* 119, no. 1 (2004): 249-275.

Cameron, Colin A., and Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press, 2005.

Cutler, D., and G. Miller. “The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.” *Demography* 42, no.1 (2005): 1–22.

Czaplicki, Alan. “Pure Milk is Better than Purified Milk Pasteurization and Milk Purity in Chicago, 1908–1916.” *Social Science History* 31, no.3 (2007): 411–33.

Dahlberg, Arthur C., and Harold S. Adams “Sanitary Milk and Ice Cream Legislation in the United States—A Study of Laws and Ordinances Establishing Sanitary Standards for Milk, Cream, and Ice Cream.” National Research Council, Washington, DC: National Academy of Sciences, 1950.

Davis, V. Effa, “Breast Feeding.” Bulletin: Chicago School of Sanitary Instruction 13 (June 8, 1910): 2.

Dube, Arindrajit, and Ben Zipperer. “Pooled Synthetic Control Estimates for Recurring Treatments: An Application to Minimum Wage Case Studies.” Unpublished paper, 2013.

Ericson, John. “Chlorination of Chicago’s Water Supply.” *American Journal of Public Health* 8, no. 10 (1918): 249–51.

Ferrie, J. P., and W. Troesken. “Water and Chicago’s Mortality Transition, 1850–1925.” *Explorations in Economic History* 45, no. 1 (2008): 1–16.

Fogel, Robert W. “Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy1.” *The American Economic Review* 84.3 (1994): 369-395.

Frank, Leslie C., and Frederic J Moss. “The Extent of Pasteurization and Tuberculin Testing in American Cities of 10,000 Population and Over in 1927 and 1931. US Public Health Service, 1931

Haines, Michael R. “Inequality and Childhood Mortality: A Comparison of England and Wales, 1911, and the United States, 1900.” *The Journal of Economic History* 45, no. 04 (1985): 885–912.

------. “American Fertility in Transition: New Estimates of Birth Rates in the United States, 1900–1910.” *Demography* 26, no.1 (1989): 137–48.

------. “Inequality and Infant and Childhood Mortality in the United States in the Twentieth Century.” *Explorations in Economic History* 48, no. 3 (2011): 418–28.

Hall, Carl W., and M. G. Trout. *Milk Pasteurization*. Westport, CT: AVI Pub. Co, 1968.

Hansen, Bruce E., *Econometrics* [Online]. June 2015 Available at: http://www.ssc.wisc.edu/~bhansen/econometrics/.

Hastings, Justine S. “Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California.” *American Economic Review* 94.1 (2004): 317-328.

Jenkins, Jane E. “Region, Politics, Pasteurization, and the Naturalizing Myth of Pure Milk in 1920s Saint John, New Brunswick.” *Acadiensis* 37, no.2 (2008): 86–105.

Kelly, Ernest, and Clarence E Clement. *Market Milk*. New York: J. Wiley & Sons, Inc., 1931.

King, G., and L. Zeng. “The Dangers of Extreme Counterfactuals.” *Political Analysis* 14, no. 2 (2006): 131–59.

Lee, Clive H. “Regional Inequalities in Infant Mortality in Britain, 1861–1971: Patterns and Hypotheses.” *Population Studies* 45, no. 1 (1991): 55–65.

Lee, K. S. “Infant Mortality Decline in the Late 19th and Early 20th Centuries: The Role of Market Milk.” *Perspectives in Biology and Medicine* 50 no.2 (2007): 585–602.

McKeown, T. *The Modern Rise of Population*. New York: Academic Press, 1976.

Meckel, R. A. *Save the Babies: American Public Health Reform and the Prevention of Infant Mortality, 1850–1929*. Johns Hopkins University Press, 1990.

Miller, Grant. “Women’s Suffrage, Political Responsiveness, and Child Survival in American History.” *The Quarterly Journal of Economics* 123, no. 3 (2008): 1287.

Millward, Robert, and Frances N. Bell. “Economic Factors in the Decline of Mortality in Late Nineteenth Century Britain.” *European Review of Economic History* 2, no. 3 (1998): 263–88.

Olmstead, Alan L., and Paul W. Rhode. “An Impossible Undertaking: The Eradication of Bovine Tuberculosis in the United States.” *The Journal of Economic History* 64.03 (2004): 734–72.

------. “Not on my Farm! Resistance to Bovine Tuberculosis Eradication in the United States.” *The Journal of Economic History* 67.03 (2007): 768–809.

Phillips, Jim, and Michael French. “State Regulation and the Hazards of Milk, 1900–1939.” *Social History of Medicine* 12, no.3 (1999): 371–88.

Preston, S. H., and M. R. Haines. *Fatal Years: Child Mortality in Late Nineteenth-Century America*. Princeton, N.J.: Princeton University Press, 1991.

Roberston, J. D. *Annual Report of the Department of Health of the City of Chicago for the Years 1911–1918 Inclusive*. Chicago, IL: Chicago. Department of Health, 1919.

Selitzer, Ralph. *The Dairy Industry in America*. New York: Magazines for Industry, 1976.

Vögele, Jörg, and Wolfgang Woelk. “Public Health and the Development of Infant Mortality in Germany, 1875–1930.” *The History of the Family* 7, no.4 (2002): 585–99.

Wagstaff, Adam. “Estimating health insurance impacts under unobserved heterogeneity: the case of Vietnam's health care fund for the poor.” Health economics 19.2 (2010): 189-208. *Winslow, C-E A. Man and Epidemics. Princeton, NJ: Princeton University Press, 1952.*

Wolf, J. H. “Low Breastfeeding Rates and Public Health in the United States.” *American Journal of Public Health* 93, no. 12 (2003): 2000–10.

Wolfers, Justin. “Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results.” *American Economic Review* 96.5 (2006): 1802-1820.

CHAPTER II.

THE PUBLIC HEALTH EFFECT OF MANDATORY MILK PASTEURIZATION ORDINANCES IN THE UNITED STATES: SYNTHETIC CONTROL METHODS APPLIED TO MULTIPLE INSTANCES

Huiqiang Wang

Department of Environmental and Natural Resource Economics,

University of Rhode Island, Kingston, RI, USA 02881

July 2015

Abstract

This study examines causal health effect of mandatory city pasteurization ordinances in the United States. I apply the synthetic control methods to multiple treated units (MTSCM). Results indicate noticeable health benefits are observed in some cities but not all. For inferences, non-parametric rank-sum tests are preferred because of non-normal outcomes in the control group. This study also suggests regression based Difference-in-Difference (DD) models lead to different results than SCM, since SCM reveals more information like unit-varying and time-varying treatment effect. This study provides an example of how SCM could supplement DD methods in practice.

Keywords: Public Health, Pasteurization, Synthetic Control Methods; Difference-in-Difference

1. INTRODUCTION

Focusing on mandatory food safety measures, this study uses case studies to illustrate causal effects estimation with multiple treated units and proper counterfactuals. For mandatory food safety policies, the mandatory framework has been well established in prior literatures, for example Henson and Caswell (1999), Segerson (1999), Fares and Rouviere (2010). However, persuasive quantitative studies are still rare. In particular, health related policies studies are not enough to meet the increasing concerns for food safety crises in recent years.

There are two challenges for researchers in this kind of study. First, researchers need a proper empirical strategy to set-up a causal relationship between the target policy and outcomes of interest. Causality directions have drawn substantial concerns from modern economics and econometrics. From Since the classic simultaneous equations models of the Cowles Commission, researchers have defined different approaches to discuss causality in economics (Hoover 2008). In econometrics, important causality concerns include distinguishing exogenous and endogenous variables, setting up conditions for identifiable causal relationships and making valid inferences for causal parameters (Cameron and Trevidi 2005). Second, from a policy perspective, researchers also need to find an appropriate way of interpreting empirical evidence. As Rodrik (2008) highlighted, to evaluate economic policies, researchers need both unit-specific and cross-sectional evidence. Cross-sectional regression results without support from unit specific case studies may be invalid because unit specific values have been “over-simplified”. Similarly, unit-specific evidence also needs to be supported by cross-sectional results for a proper economic interpretation. In recent

years, efforts to combine both unit-specific and cross-sectional evidence are increasing as more econometric tools become available.

In this study, I apply synthetic control methods (SCM) to multiple treated units and measure causal health benefits of mandatory milk pasteurization ordinances in 1916. As a food safety innovation, pasteurization was believed as a key measure to control milk diseases in history. In the late 1800s, biological contaminations caused serious milk diseases (Figure 1). Historians described them as dangerous as "White Plague" (Seltzer 1976). Starting in the 1890s, pasteurized milk was provided to the public in city milk depots. Early experiments recorded remarkable health benefits of milk pasteurization, particularly the drop in childhood diarrhea mortality rates (Kelly and Clement 1931). In addition, pasteurization helped to control other milk epidemics, like typhoid fever and scarlet fever. Medical professionals recommended it as "practically feasible to keep milk clean and pure" and an "essential safeguard" (Winslow 1952). In addition, pasteurization was preferable for large-scale liquid milk production. The principle of pasteurization is to eliminate pathogens at some temperature that will not alter the physical and nutritional attributes of milk (Hall and Trout 1968).

As a remarkable public health innovation, milk pasteurization was regarded a key in the fight of milk diseases in history. In the middle 1890s, biological contamination of milk was a serious health threat, particularly to children (Figure 1). In some cities, pasteurization was applied to clean milk on a voluntary basis as early as the late 1800s. The technology was able to eliminate almost all pathogens at temperatures that avoid physical changes and nutritional losses (Kelly and Clement

1931; Hall and Trout 1968). In addition, this innovation was cost efficient for large scale commercial production. Early case studies indicated the health consequences of pasteurization were amazing, and starting in the early 1910s, some cities began to require milk to be pasteurized before sale.

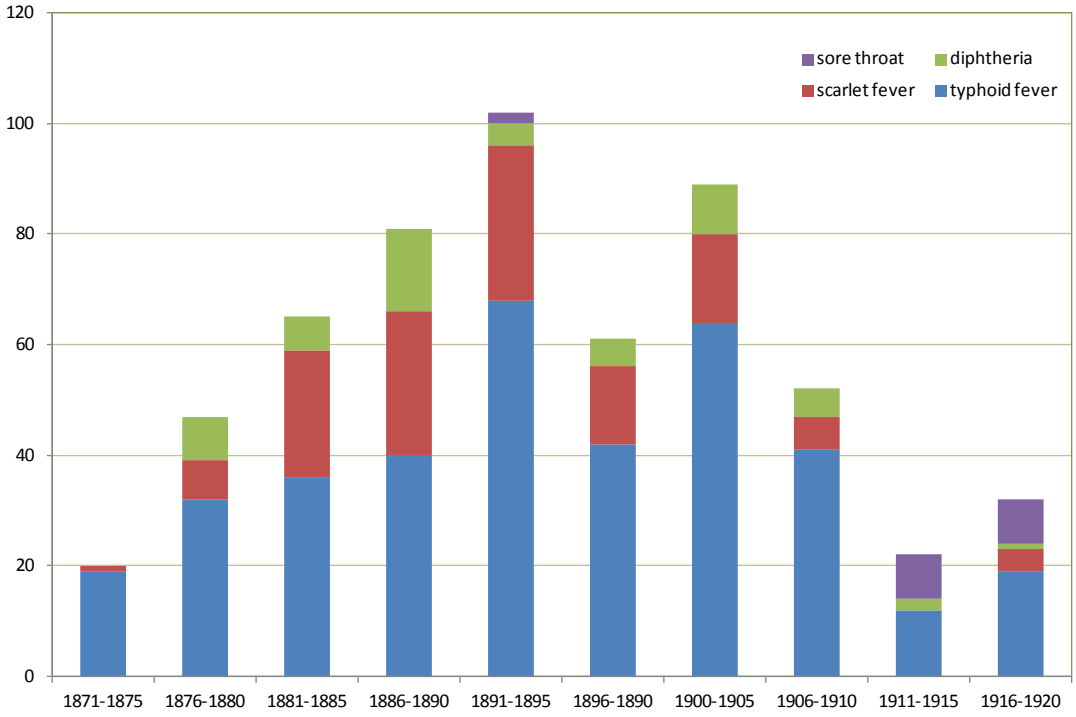
For example, Chicago's child diarrhea mortality increased in the early 1900s with a rise in bottle-feeding involving contaminated milk. It was the introduction of the full milk pasteurization ordinance that caused a structural change in child diarrhea mortality rates. On the other hand, this innovation was controversial since there were complex tradeoffs between interests of dairy farmers, milk consumers and city health officials (Czaplicki 2007). Similarly stories also occurred in other cities (Levitt 1996). Thus, it has been an interesting policy question to know the causal health impact of milk pasteurization ordinances, *e.g*. whether they were "large-scale public health innovations" in the early 20th century in the United States (Cutler and Miller 2005). Unlike prior narrative studies, this paper aims to provide a clear causal estimation for milk ordinances in Chicago and other five cities which adopted mandatory ordinances in 1916. These cities are chosen for cross-sectional comparison because of the consistent timing of their interventions.

I apply synthetic control methods (SCM) as the empirical approach to estimate the causal relationship. SCM was introduced by Abadie and Gardeazabal (2003) and became mathematically formalized in Abadie, Diamond and Hainemuller (ADH 2010, 2014). SCM is more than a bridge between quantitative and qualitative studies. It also connects unit-specific and cross-sectional evidences (Billmeier and Nannicini 2013). In this study, regular SCM algorithm is extended to multiple treated units (MTSCM).

The purposes of this paper are three-fold. First, it measures causal health effects of pasteurization ordinances and makes valid inferences. Second, as an extension of SCM, this study discusses how to make valid causal inferences with multiple treated units. Third, this study also compares performance of SCM and Difference-in-Difference (DD) in practice. SCM might be a supplement for DD applications (ADH 2010), but robust analyses of their estimates and related inference problems are still lacking. This study aims to fill that void.

This paper proceeds as follows. Part 2 provides a background review. Part 3 introduces the data used in this analysis. Then, MTSCM estimations and inferences are made in Part 4. Part 5 illustrates differences between DD and MTSCM. Finally, concluding remarks are made in Section 6.

Figure 2- 1 Number of Reported Milk Diseases Epidemics 1871 to 1920 (every 5 years)



Source: the original annual data are obtained from North (1921) and summed up every five years by the authors. The data did not include all epidemics in observed periods (more in North, 1921).

2. BACKGROUND

2.1 Milk Diseases and Pasteurization

From the late 19th to early 20th century, a safe milk supply was thought to be a key health innovation in cities. In the second half of the 19th century, the market milk supply was riddled with intentional adulterations and biological contaminations. Lower quality milk was the source of many epidemics, including diarrhea, tuberculosis, scarlet fever and sore throat. These diseases were particularly dangerous to children after females joined the labor force and increasingly relied on bottle feedings. For example, the study of Wolf (2003) illustrated how unclean milk, bottle feeding and high incidence of child diarrhea deaths were correlated in Chicago in the early 1900s. Outside the United States, researchers also noticed the co-movements of children health and improved milk quality in the United Kingdom and Germany (Beaver 1973; Meckel 1990; Vögele and Woelk 2002).

At that time, one solution for milk problems was pasteurization. Commercial milk pasteurization started from city milk depots sponsored by philanthropists (Selitzer 1976). Later, milk dealers also benefited from this innovation, as milk can be preserved longer after heating, so voluntary adoptions became increasing popular. For example, the share of pasteurization rose from 5% to 40% in New York from 1902 to 1912 (Jordan 1913), though the extent of pasteurization was still quite low in many cities which had no formal requirements. Table 2 shows that the averaged extent of pasteurization was less than 70% in those cities, even in the early 1930s. Cities with formal requirements had much more milk pasteurized.

Compared with commercial applications, mandatory pasteurization ordinances came later. The first city ordinance was adopted in Chicago in 1908, but it was then banned by the state of Illinois and its full adoption did not come until 1916. Interestingly, deaths from typhoid fever, one of the leading epidemics, declined following water purification measures, as discussed in Chapter 1. However, the structural change in children diarrhea deaths in Chicago coincided with its mandatory pasteurization ordinance. More than Chicago, Cleveland, Milwaukee, Indianapolis, Richmond and San Francisco also adopted similar milk ordinances which required all milk pasteurized before sale. [48]

Similar to some modern food innovations, pasteurization also met strong opposition. Some are health concerns, for example possible nutrient loss, physical and flavor changes of pasteurized milk, and long-term health impacts to children (Hall and Trout 1968). On the other side, the issue was complicated by interest conflicts between dairy farmers and city health officials. When bovine tuberculin tests and pasteurization became mandatory, farmers raised strong resistance to these regulations (Olmstead, and Rhode 2004). However, from the perspective of overloaded city health officials, pasteurization was an economical and efficient policy tool to control milk diseases at that time (Czaplicki 2007). From a modern perspective, the key to better understand this debate is understanding the causal health impact of this policy intervention historically. If pasteurization ordinances were significantly responsible for transitions in health outcomes, they should be given credit. Otherwise, alternatives like a combination of both mandatory and voluntary measures could be more desirable.

[48] More exactly, most ordinances required all milk but certified needs to be pasteurized before sale. But the share of certified milk was quite small in total milk supply. See Block (1999; 2009).

Table 2- 1 Sources of Information about City Milk Pasteurization Ordinance in 1916

City	State	Sources
Chicago	Illinois	Illinois Health News, Illinois State Department of Health, 1922, p. 144-145
Cleveland	Ohio	US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1915, p. 217-224
Indianapolis	Indiana	Journal of the Indiana State Medical Association, 1916, Feb., p.71; US Public Health
Milwaukee	Wisconsin	Hibbard B. and Erdmann H., Marketing Wisconsin milk, 1917, p. 49-50; Levitt J. W., The Healthiest City: Milwaukee and the Politics of Health Reform, 1996, p. 187
Richmond	Virginia	US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1915, p. 364-365
San Francisco	California	US Public Health Service, Reprint from the Public Health Reports, 1916, p. 160-173

2.2 Causal Effect Estimators

As an extension of SCM, MTSCM has been used in estimating policy effects in the multiple treated units in recent years, for example Billmeier and Nannicini (2013), Gobillon and Magnac (2013); and Dube and Zipperer (2014). MTSCM can be described as a two-step process. First, standard single treated unit SCM can be used to generate outcomes of interest, usually as real-synthetic gaps. Second, all outcomes are collected for causal inference, using either parametric or non-parametric techniques.

Compared with SCM, MTSCM has some specific concerns. For example, researchers should be careful in choosing a proper method for MTSCM inferences. In SCM, non-parametric permutation or “placebo studies” are generally used for the significant tests with only one treated unit. When more than one treated units are available, researchers then have more tools for inferences. For example, one can use one sample or two samples t-tests if outcomes are normally distributed. If normality assumption cannot be satisfied, non-parametric techniques should be used. Thus, one

contribution of this study is to discuss how to make valid inferences with small size non-normal samples in MTSCM applications.

Another specific feature of MTSCM is its connection with DD models. With multiple units exposed to intervention, I can conduct a DD regression using the MTSCM panel. Then, I will check how MTSCM reveals information which cannot be reflected in DD. In principle, MTSCM differs from DD with the weights used in constructing the counterfactual. DD models measure "population average difference" (Imbens and Wooldridge 2007), but MTSCM measures "population optimized difference" via an explicit weight selection algorithm, which generates non-negative weights summing to one. In addition, SCM reveals more information than DD. MTSCM can show us both unit-varying and time-varying treatment effect which cannot be reflected by DD.

As mentioned above, this paper focuses on how marginal changes in pasteurization (p) were responsible for changes in the conditional mean of a function of health outcomes, holding other explanatory variables ($\mathbf{X}$) constant, as (1). [49]

$$\begin{aligned} &\text{IF } h(p,\mathbf{X}) = \mathbb{E}\left(Y \mid p,\mathbf{X}\right) \text{ function of heath effects} \\ &\nabla_p h(p,\mathbf{X}) = \begin{cases} \dfrac{\partial}{\partial p} h(p,\mathbf{X}) & \text{with a continuous } p \\ h(p=1,\mathbf{X}) - h(p=0,\mathbf{X}) & \text{with a binary } p \end{cases} \end{aligned} \tag{1}$$

Here, the effect of pasteurization can be either specified as a binary policy intervention or as a continuous variable to proxy the change in the share of milk that is pasteurized. Herein, pasteurization ordinances are considered as discrete binary variables.

[49] One important but not explicitly explained point in the set-up of SCM (ADH 2010, 2014) is the issue of endogeneity. SCM has no assumption about the endogenous bias, which should be a concern for this method. In this study, however, historical evidence suggested mot pasteurization ordinances were not adopted to control children diarrhea mortality. An important fact was early ordinances were used for to eradicate bovine tuberculosis. Thus, the endogeneity assumption can be relaxed in this study.

3. DATA

The study period is from 1900 to 1930, a time which witnessed a rapid expansion of pasteurization in the United States. The dependent variable is the annual city-level diarrhea and enteritis mortality rate for children under 2 years,[50] calculated as the number of deaths in every 1,000 children under 5 years of age. Annual population data are calculated with arithmetic averages of census data from 1900, 1910, 1920 and 1930. The number of diarrhea deaths is obtained from annual Mortality Statistics 1900 to 1931. The year 1900 is the first year I can obtain city mortality statistics in the United States. The 1931 survey had records of the extent of pasteurization and the *status quo* of city ordinances. So the year 1931 is a good ending point in this study.

The treatment group includes 6 cities: Chicago, Cleveland, Indianapolis, Milwaukee, Richmond and San Francisco, all of which adopted pasteurization ordinances in 1916. For the control group, there are 52 cities which had no mandatory ordinances by 1931 (Frank and Moss, 1931). Compared with the treated units, the extent of pasteurization in the control units was lower, as mentioned. In 1931, the averaged share of pasteurization was 99.2% in the treated group, while it was only 65.6% in the control group (Table 2).

In DD, observed covariates are used to solve possible selection bias associated with the intervention variable. Similarly, SCM also requires a set of predictors to construct the counterfactual. The predictors help to select control units which are close to the treated units in non-outcome covariates. Moreover, lagged dependent variables are usually added as covariates in DD models to capture dynamic trends and control

[50] Diarrhea death is a good indicator to milk quality (Lee 2007). In addition, prior studies like Cutler and Miller (2005) have not discussed the effect of major public health campaign on this variable.

autocorrelations in unobserved dependence. For SCM, pre-intervention outcomes are similarly important.[51] Selection of control units thus requires both a set of proper predictors and enough pre-treatment observations.

In this study, non-outcome covariates for DD and SCM include city population, demographic characteristics, and income. These predictors include total population (log values), average wage in manufacture, female share, white share, share of population under 5 years. Population and demographic variables are obtained from the decennial Census of Population years 1900, 1910, 1920 and 1930. Income information is from the Census of Manufactures years 1900, 1909, 1919 and 1929.

[51] More discussion can be found in Dube and Zipperer (2014).

Table 2- 2 Share (%) of pasteurized milk in cities 1931

Treated	%	Control	%	Control	%	Control	%
		Brockton	84.0	Memphis	74.0	Quincy	70.0
		Concord	28.0	Meriden	75.0	Raleigh	33.3
Chicago	99.5	Duluth	58.0	Middletown	88.0	Sacramento	89.5
		Evansville	85.0	Mobile	15.0	San Antonio	69.0
Cleveland	99.0	Fitchburg	66.2	Montclair	82.3	San Diego	75.5
		Green Bay	80.0	Muncie	75.0	Savannah	33.0
Indianapolis	97.5	Hartford	89.0	Nashville	60.0	Seattle	87.9
		Jackson City	58.0	New Britain	68.0	Springfield	85.0
Milwaukee	99.5	Jacksonville	40.0	New Haven	80.0	St. Paul	79.7
		Jamestown	25.0	New Orleans	----	Superior	41.0
Richmond	100	Kalamazoo	84.0	Omaha	70.0	Tacoma	54.0
		Kansas city	50.0	Paducah	60.0	Troy	39.5
San Francisco	99.5	La Fayette	35.0	Petersburg	64.0	Utica	79.9
		Lancaster	70.0	Pittsfield	64.0	Wheeling	76.0
		Lincoln	80.0	Plainfield	71.0	Wichita	66.0
		Los Angeles	82.3	Portland (ME)	86.7	Wilmington	40.0
		Lynchburg	33.3	Portland (OR)	75.0		
Avg. treated	99.2	Manchester	85.0	Providence	86.9	Avg. control	65.6

Source: Frank and Moss, *The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931*. US Public Health Service

4. SYNTHETIC CONTROL ANALYSIS

4.1 Specification

SCM was originally designed for comparative case studies, *i.e.* with only 1 treated unit. As mentioned above, SCM is extended to multiple treated units as MTSCM in this study. In the MTSCM setup, treatment effect estimation of each treated unit follows a standard SCM algorithm. Equation (2) illustrates how optimized weights are generated in SCM. First, suppose I have *M* treated cities, with each city *i* (*i* = 1, 2 … *M*) having a pasteurization ordinance, and *N* control cities *j* (*j* = 1, 2 … *N*) without such ordinances. Interventions split the study period into pre-treatment period *T* and post-treatment *t*. Thus, Y_{iT} and Y_{jT} are pre-treatment outcomes of *T*. Similarly, Y_{it} and Y_{jt} are the post-treatment results in *t*. **X** is a vector covariates. For a single treated unit *i*, the treatment effect Φ_{it} is measured as the difference between its real value and a convex combination of its control units with optimized weights w_j^*, as (2). Roughly speaking, the difference between the real treated and its counterfactual (real-synthetic gap) are akin to the treatment effect on the treated in a linear framework.

As mentioned, the optimized weight is driven by minimizing the distance δ_m between $\mathbf{X}_\mathrm{i}$ and $\mathbf{X}_\mathrm{i} \cdot \mathbf{W}_\mathrm{j}$ in the pre-intervention period T[52], as in Equation (2). With the choice of matrix **V**, I then minimize different distances (3).

$$\Phi_{it} = Y_{it} - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \quad \text{with s.t. } w_j^* \geq 0 \text{ and } \sum_{j=1}^{N} w_j^* = 1 \tag{2}$$

$$\delta_m = \min_w [(\mathbf{X}_i - \mathbf{X}_j \cdot \mathbf{W})' \mathrm{V} (\mathbf{X}_i - \mathbf{X}_j \cdot \mathbf{W})]^{1/2} \tag{3}$$

[52] Abadie, Diamond and Hainmueller, "Synthetic Control Method for Comparative Case Study" p.496

Alternatively, w_j^* can be obtained by minimizing pre-intervention real-synthetic gaps if dependent variables are auto-correlated. However, the optimized weights may be different if we use another set of predictors. Computations of w_j^* are finished via a non-parametric algorithm as (4).

$$Y_{iT} - \sum_{j=1}^{N} w_j^* \cdot Y_{jT} \simeq 0 \text{ with s.t. } w_j^* \geq 0 \text{ and } \sum_{j=1}^{N} w_j^* = 1 \tag{4}$$

Equation (5) below shows how the treatment effect is measured in MTSCM when there is more than one unit exposed to interventions. MTSCM allocates optimized weights to the comparison units to make a combination of them sufficiently close to the treated unit. In SCM, $Y_{it}(0)$ becomes "observable", given $D_i = 1$.

$$\begin{aligned} \Phi^{SCM} &= \mathbb{E}[Y_{it} - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \mid \mathbf{X}_i, \mathbf{X}_j] \\ &= \mathbb{E}[Y_{it} \mid \mathbf{X}_i] - \mathbb{E}[\sum_{j=1}^{N} w_j^* \cdot Y_{jt} \mid \mathbf{X}_j] \\ &\simeq \mathbb{E}[Y_{it} - Y_{it} \mid \mathbf{X}_i \ ] \\ \text{IF } Y_{it}(0) &- \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \simeq 0 \end{aligned} \tag{5}$$

As mentioned, this study also concerns performances of SCM and DD. ADH (2010) suggested SCM could be used as a supplement of DD. This point is formally expressed as Equation (6) below, which shows how DD and MTSCM are connected. MTSCM coefficients are based on the optimized w_j^*. Instead, DD uses averaged weight w, where $w = 1/N$. So DD can be regarded as a special class of MTSCM, when $w_j^* = w = 1/N$.

$$
\begin{aligned}
\Phi^{DD} &= \mathbb{E}[Y_{it} - Y_{jt} \mid \mathbf{X}_i, \mathbf{X}_j] - \mathbb{E}[Y_{iT} - Y_{jT} \mid \mathbf{X}_i, \mathbf{X}_j] \\
&= \mathbb{E}[Y_{it} - \sum_{j=1}^{N} \frac{1}{N} \cdot Y_{jt} \mid \mathbf{X}_i, \mathbf{X}_j] - \mathbb{E}[Y_{iT} - \sum_{j=1}^{N} \frac{1}{N} \cdot Y_{jT} \mid \mathbf{X}_i, \mathbf{X}_j] \\
&= \mathbb{E}[Y_{it} - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \mid \mathbf{X}_i, \mathbf{X}_j] \qquad (6) \\
&= \Phi^{SCM} \\
\text{IF } w_j^* &= \frac{1}{N}; \text{ and SCM makes } Y_{iT} - \sum_{j=1}^{N} w_j^* \cdot Y_{jT} \simeq 0
\end{aligned}
$$

Using sample data, the averaged MTSCM treatment effect as (7) when we have *N* control cities and *M* treated cities. MTSCM measures the averaged real-synthetic gaps at period *t*.

$$
\hat{\Phi}^{SCM} = \frac{1}{M} \cdot \sum_{i=1}^{M} [Y_{it} - \sum_{j=1}^{N} w_j^* \cdot Y_{jt}] \qquad (7)
$$

In this study, I split the MTSCM process into three steps, namely building synthetic versions, implement placebo studies, and make statistical inferences.

(1) SCM is applied to six treated units. Each synthetic version is constructed with the same 52 control cities. Predictors include demographic variables, income and lagged dependent outcomes, as mentioned above. If ordinances were effective, the real mortality trajectory should be lower than its synthetic version. In other words, there should be a “negative” real-synthetic gap.

(2) Then, I make placebo studies for all 52 control cities using the same SCM algorithm. If the treatment effect was not from random chance, the effect should be more noticeable in the treated cities. For each control city, its synthetic version is constructed from the other 51 control units only.

(3) I conduct statistical inference with the post-treatment real-synthetic gaps in the treated and control groups. For multiple treated units, causal inferences can be made in different ways, depending on the properties of the outcome distribution. If the sample is normal, we can use a t-test comparing sample means. Otherwise, non-parametric methods are more preferable, for example Wilcoxon rank-sum tests or Mann Whitney U statistics.

4.2 Results

At the individual level, real and synthetic trajectories of treated cities are depicted in Figure 2. Their SCM weights are reported in Table 3. We see that SCM weights are obviously different from the averaged weight, as 1/52 or 0.019. In Table 3, some weights are zero, while some are larger than the averaged value. The real-synthetic gaps should be negative if the treatment was effective (the real trajectory should be lower).

One noticeable feature in Figure 2 is treatment effects vary across treated units and over periods. As Abadie (online) noted treatment effects may not be observed immediately after interventions, so it is recommended to include enough post-intervention observations for the treatment to be observable. In Chicago and Cleveland, noticeable treatment effects were observed after 1920. Richmond and San Francisco also have some real-synthetic gaps, but their real and synthetic trajectories diverged before 1916 so it is uncertain whether the gaps were the result of a causal intervention effect or just a lack of fitting before intervention. In contrast to Chicago and Cleveland, expected negative real-synthetic gaps did not appear in Indianapolis

and Milwaukee. In Milwaukee, the gaps were small and fluctuated a lot in post-intervention periods. In Indianapolis, real mortality rates were higher than synthetic values in most periods but the real mortality rate was on a faster declining trend in the 1920s and there was essentially zero gap by 1930. In sum, the results suggest that intervention effects were not quite consistent among treated units. Substantial treatment effects existed only in some cities.

Figure 2- 2 Mortality rate trend real treated cities and their synthetic versions (Treated-blue solid line; Control-red dash line)

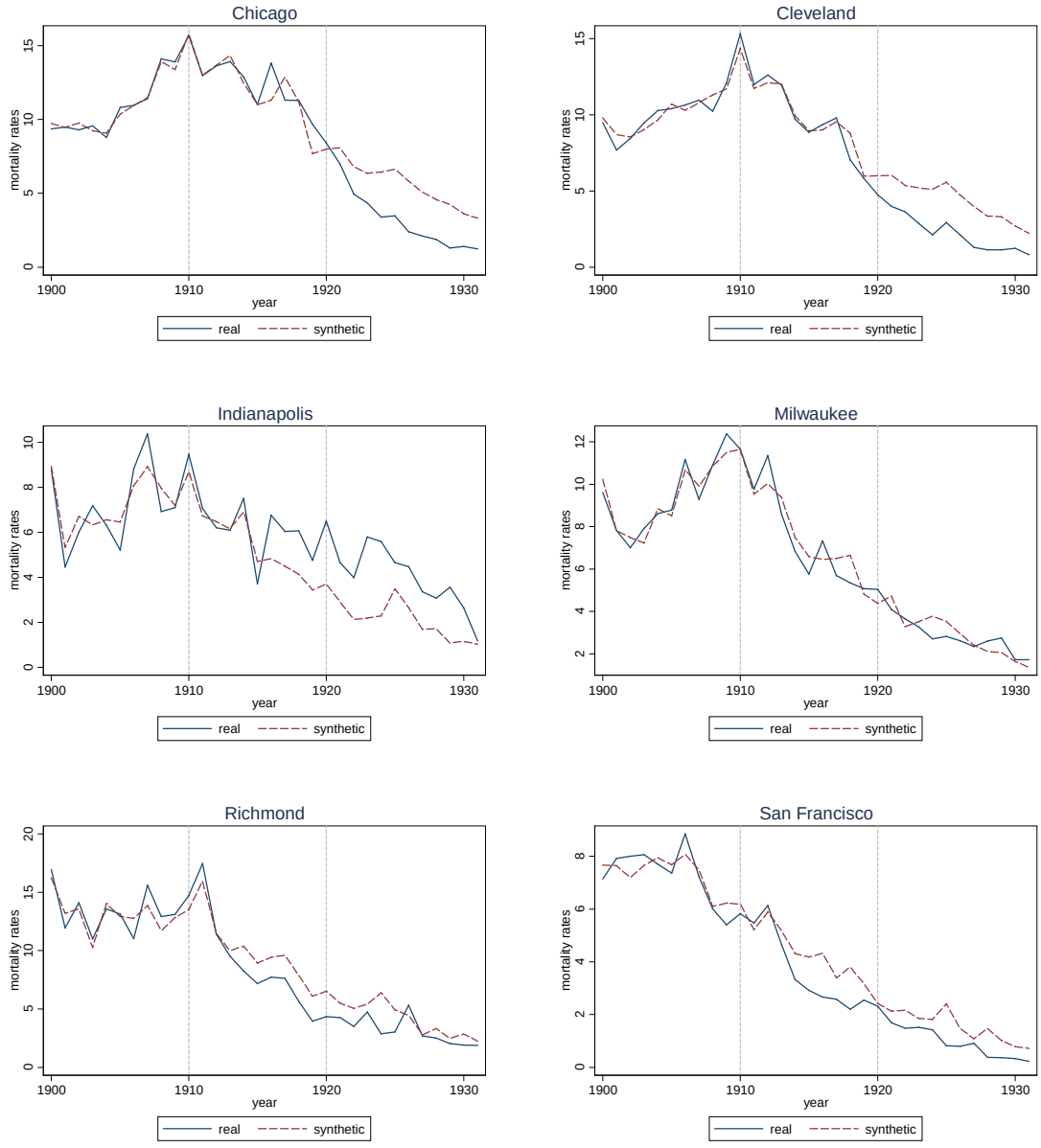


Table 2- 3 SCM Weights of Treated Cities, 52 Controls

Treated / Control	CHI	CLV	INP	MIK	RMD	SFC
Brockton	0	0	0	0	0	0
Concord	0	0	0	0	0	0.111
Duluth	0	0	0	0.051	0	0
Evansville	0	0	0	0	0	0
Fitchburg	0	0.047	0	0.107	0	0
Green Bay	0	0.022	0	0	0	0
Hartford	0	0.035	0	0	0	0
Jackson City	0	0	0	0	0	0
Jacksonville	0	0	0	0	0	0
Jamestown	0	0	0	0	0	0
Kalamazoo	0	0.057	0	0	0	0
Kansas city	0	0	0	0	0	0
La Fayette	0	0.289	0.058	0	0	0
Lancaster	0	0.091	0	0	0	0
Lincoln	0	0	0.342	0	0	0
Los Angeles	0	0	0	0	0	0
Lynchburg	0.034	0	0	0	0	0
Manchester	0	0	0	0	0.139	0
Memphis	0	0	0.034	0	0.097	0.046
Meriden	0	0	0.054	0	0	0
Middletown	0	0	0	0	0	0
Mobile	0.103	0	0	0	0	0
Montclair	0.013	0.231	0	0.009	0	0
Muncie	0	0	0.102	0	0	0
Nashville	0	0	0	0.055	0.357	0
New Britain	0	0	0.075	0	0.004	0

Note: Chicago (CHI); Cleveland (CLV); Indianapolis (INP); Milwaukee (MIK); Richmond (RMD); San Francisco (SFC);

Table 2- 3 Continued

Treated / Control	CHI	CLV	INP	MIK	RMD	SFC
New Haven	0	0	0.075	0	0.004	0
New Orleans	0	0	0	0	0	0.122
Omaha	0	0	0	0	0.033	0
Paducah	0	0	0.199	0	0	0
Petersburg	0	0	0	0	0	0.097
Pittsfield	0.078	0.047	0	0.065	0	0
Plainfield	0	0.022	0	0	0	0
Portland	0.041	0.035	0	0.018	0	0.171
Portland	0	0	0	0	0.153	0
Providence	0	0	0	0.326	0	0.027
Quincy	0	0	0	0	0	0
Raleigh	0.024	0.057	0.052	0	0	0
Sacramento	0	0	0	0	0.169	0
San Antonio	0.301	0.289	0	0.082	0	0
San Diego	0.143	0.091	0	0	0	0
Savannah	0	0	0	0	0	0
Seattle	0.056	0	0	0	0	0
Springfield	0	0	0	0	0	0.425
St. Paul	0	0	0	0	0	0
Superior	0	0	0	0	0	0
Tacoma	0	0	0	0	0	0
Troy	0	0	0	0	0	0
Utica	0	0	0	0	0	0
Wheeling	0.206	0.231	0.084	0.233	0	0
Wichita	0	0	0	0	0	0
Wilmington	0	0	0	0	0	0.001

Note: Chicago (CHI); Cleveland (CLV); Indianapolis (INP); Milwaukee (MIK); Richmond (RMD); San Francisco (SFC);

4.3 Statistical Inferences

This section discusses valid statistical inference for the above estimates. Unlike single treated unit SCM, MTSCM have more than one unit exposed to interventions. Instead of a permutation test for a single treated unit, I need an inference technique to reveal the overall treatment effect at the group level. If two groups are of similar sizes and with Gaussian distributions of sample mean, I can use two sample *t*-tests. However, the sample sizes of the two groups are very different, 6 and 52. Considering the small size of the treated group, there are two options. First, if outcomes in the control group are close to a normal distribution, I can use a one-sample *t*-test to compare the mean of the control group with the averaged value of treated units. Of course, this approach is requires a normal distribution as well, and the average treatment effect may be over-simplifying the difference among treated units. The second option is to use other non-parametric tests like Wilcoxon Rank-sum or Mann-Whitney U tests. Here I will practice the two approaches and discusses their differences, and check whether sample mean comparisons can provide us full information.

First, I conduct a one sample *t*-test. To do so, I begin with the Shapiro-Wilk (SW) normality test to see whether post-intervention real-synthetic gaps in the control group are normally distributed. [53] Sample means of treated and control groups are in Table 4. According to Shapiro and Wilk (1965), their W statistics for complete sample of normality testing can be defined as (8). In this study, if $\{Y_i\}$ are normal sample, $Y_i = \mu + \sigma X_i$.

[53] Here the real-synthetic gap is not scaled into an interval, since it is calculated as the number of deaths of population. Instead of a rate or share, this number is not limited to some lower or upper bounds.

$$W = (a'Y)/S^2 = \left(\sum_{i=1}^{n} a_i Y_i\right)^2 \Big/ \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$$
$$\text{where } S^2 = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 \tag{8}$$

Y_i is ranked from $(Y_1 \cdots Y_n)$. Coefficients $a'$is a derivation from (9).

$$a_i = (a_1 \cdots a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}} \tag{9}$$
$$\text{where } V \text{ is a n} \times \text{n covariance matrix}$$

The null hypothesis of SW is the sample $\{Y_i\}$ is normally distributed. If W statistics is lower than a threshold value, we can reject the null hypothesis, *e.g*. the sample is from a non-normal distribution.

Results of SW tests for normality of real-synthetic gaps in control group are listed in Table 4. Tests results suggest gaps of control group are normally distributed in only eight years of total 15 post-intervention observations. They are 1919, 1920, 1921, 1923, 1924, 1926, 1927 and 1931. In other years, real-synthetic gaps are non-normal. Thus, one sample *t*-tests are not valid in these years. And my SW statistics was only made in the eight year with normal outcomes. Results suggest the averaged values of treated units are significantly different from the control only in 1924, 1926, 1927 and 1931. Even using a simplified sample average, results are not consistent across post-intervention periods. In addition, eight samples do not approximate exactly normal in Figure 3, although the SW null hypothesis cannot be rejected.

Mann-Whitney U statistics (MWU, Mann and Whitney 1947) allows us to test two samples without assuming dependent variables are normally distributed. In principle, MWU test is similar to Wilcoxon rank-sum test (Wilcoxon 1945), in which the test statistics are constructed by ranking outcomes in two samples. The null

hypothesis is the two samples have the same distribution. If the null hypothesis is rejected, we can conclude the rank of one sample is significantly different from the other.

$$\begin{aligned} U_a &= R_a - n_a(n_a+1)/2 \\ U_b &= R_b - n_b(n_b+1)/2 \\ U &= \min(U_a, U_b) \end{aligned} \tag{10}$$

In Equation (10), R_a and R_b are the ranks in the two groups (a, b). Number of units in the two groups are n_a and n_b. Statistics for significance test is the smaller one of U_a and U_b. Results in Table 4 show that rank sums in treated group are not significantly different from control group, except for 1924. Overall, the treated group had no better health outcomes than their control units in each year after 1916.

In this study, one potential concern for the validity of MWU is the sample size. The small size of the treated group (*N1* = 6 and *N2* = 52) makes a very restrictive critical value of U. For robustness, I proposed an alternative power test which is based on the principle of permutation test in regular single treatment SCM inferences. It proceeds as follows. First, I calculate the sample mean in the treated group for each post-intervention period. It is the averaged real-synthetic gaps of six units in year *t*. Second, I take a random sample of six units out of total 52 units in the control group. Similar sampling is repeated M times. Third, I count how many times (M1) the absolute value of the averaged negative real-synthetic gaps in treated group is smaller than (and equal) the sample of control units ($Gap_{treated} - Gap_{control} \geq 0$). Finally, the power statistics is calculated as p = M1/M. The null hypothesis is $Gap_{treated} - Gap_{control} = 0$, e.g. there is no effect of intervention. If the power test value is smaller than the critical value (0.01 or 0.05), we reject the null hypothesis.

The sampling is repeated 1,000 times for individual years 1917-1931. Distributions of 1,000 averages of six control units are plotted in Figure 4. Table 5 reports the *p*-values calculated in each post-intervention year: the null hypothesis cannot be rejected in any years. Similar to the MWU results, real-synthetic gaps in the treatment group were not significantly different from the control group.

Figure 2- 3 Distributions of Real-Synthetic Gaps in Control Group (Selected years)

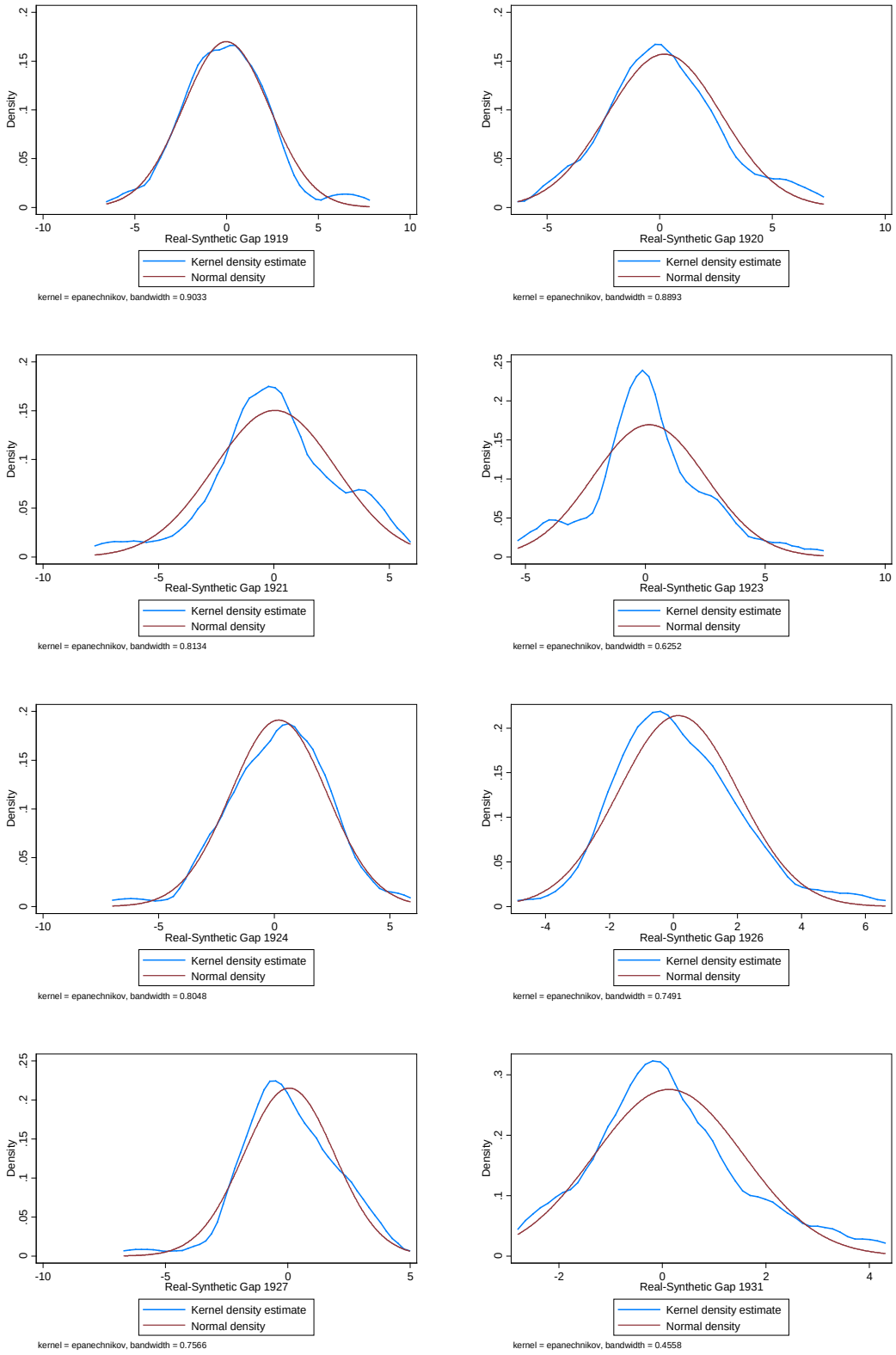


Table 2- 4 Mean Value Comparison of the Treated and Control Groups: *t*-test

Year	Treated Mean	Control Mean	Prob. > z (SW)	*p*-value (*t*-test)	Prob. > z (rank-sum)
1917	-0.3946	-0.0777	0.0000	------	1.0000
1918	-0.6125	-0.1778	0.0228	------	0.6643
1919	0.0218	-0.0151	0.3936	0.9101	0.8382
1920	0.1204	0.2131	0.6139	0.7932	0.9593
1921	-0.3635	0.0451	0.2787	0.2722	0.5570
1922	-0.6384	0.1412	0.0001	------	0.3853
1923	-0.1288	0.1702	0.2079	0.3639	0.6096
1924	-1.3911	0.1961	0.7505	0.0000	0.0555
1925	-1.1729	0.2784	0.0476	------	0.2109
1926	-0.5012	0.1573	0.2028	0.0139	0.6458
1927	-0.7489	0.0597	0.2005	0.0028	0.3715
1928	-0.8550	-0.1657	0.0095	------	0.3853
1929	-0.3169	-0.0417	0.0296	------	0.8183
1930	-0.6233	0.0290	0.0355	------	0.5068
1931	-0.7020	0.1287	0.1012	0.0001	0.1680

Note: p-value is two-tail t-test; Prob. > z (SW) is the SW statistics for normality tests; Prob. > z (rank-sum) is the differences in rank-sum of the two groups. *p*-values (*t*-test) are not reported in the year if real-synthetic gaps in the control groups were non-normal.

Table 2- 5 Power Tests of Random Sampling Group Means (N = 1,000)

Year	*p*-Value	Year	*p*-Value	Year	*p*-Value
1917	0.397	1922	0.479	1927	0.532
1918	0.420	1923	0.561	1928	0.391
1919	0.522	1924	0.571	1929	0.478
1920	0.555	1925	0.595	1930	0.505
1921	0.545	1926	0.584	1931	0.572

Figure 2- 4 Distributions of 1,000 Control Sample Averages (sample size = 6)

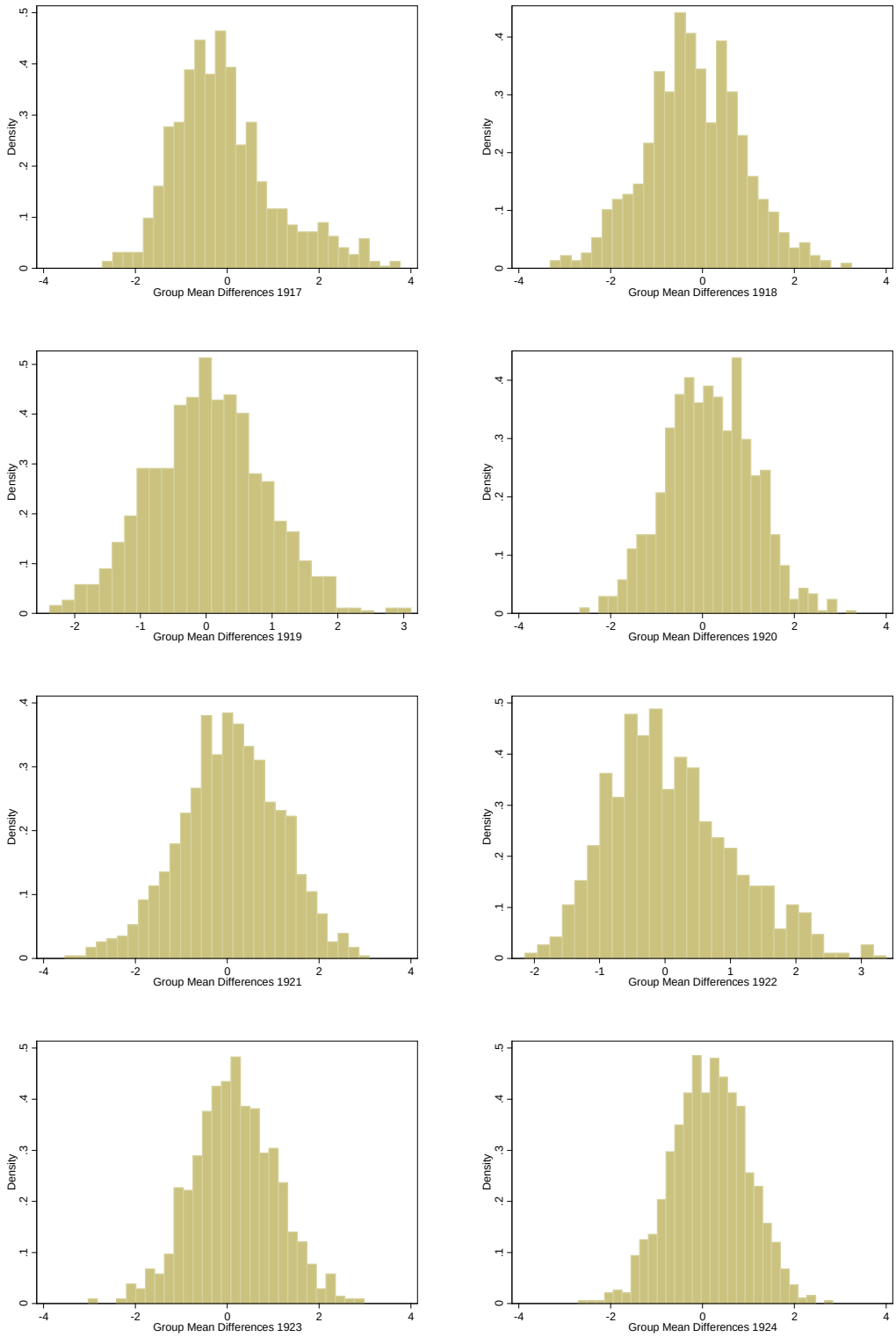
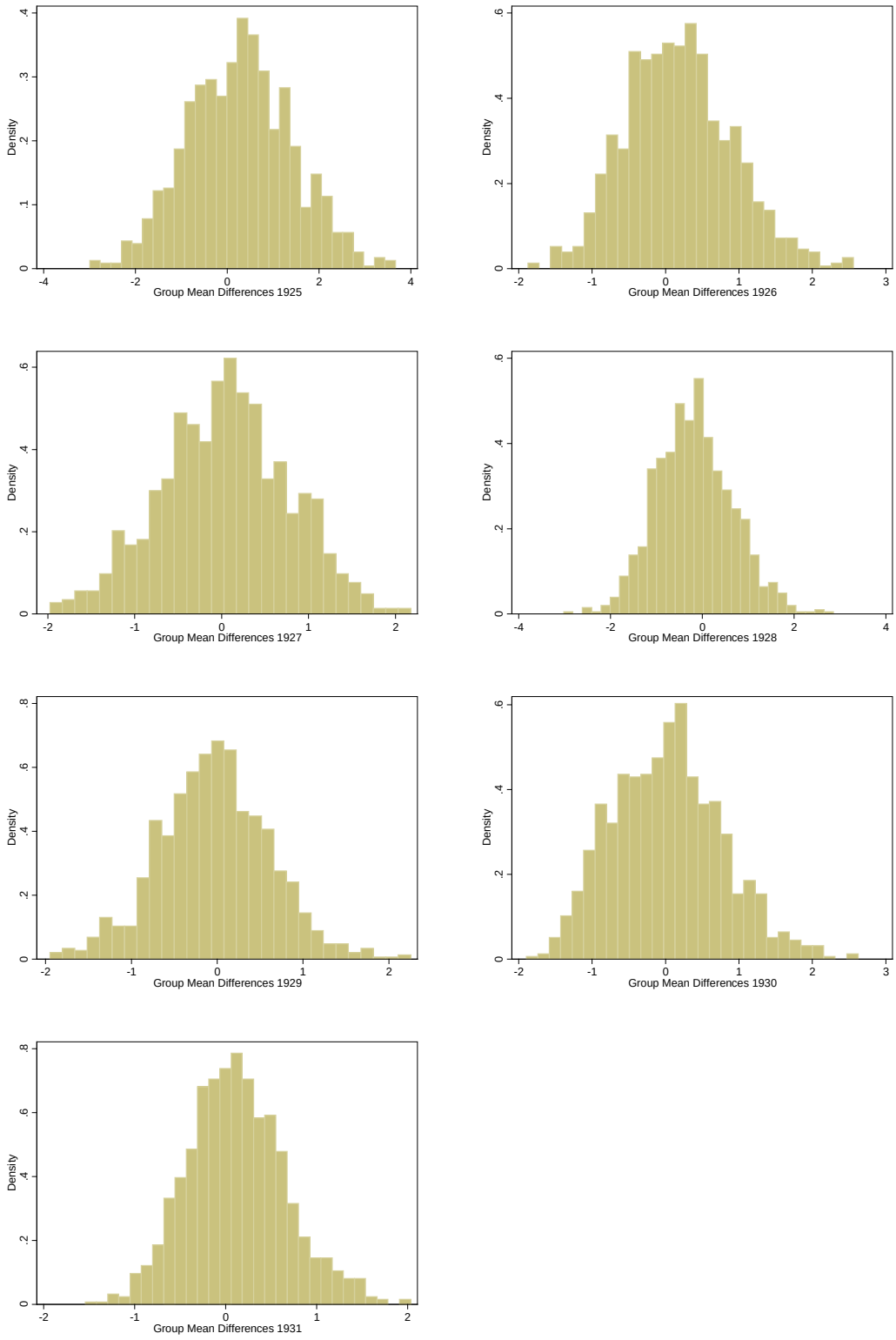


Figure 2- 4 Continued



5. DIFFERENCE-IN-DIFFERENCE MODEL

The next step is to clarify another concern, the difference between DD and SCM estimators. In Equation (6), we see how the two methods differ with conditional expectation notations. With real data, this section discusses estimation and inference with DD models and compares their performances.

The danger of serial correlation in the error term for meaningful statistical inference has been well illustrated by Bertrand *et al*. (2004). Regarding correlated errors within units, three solutions are applied in my DD specifications to obtain consistent standard errors: adding lagged terms in an autoregressive (AR) model, using clustered standard errors, and aggregating data into before and after intervention periods. The DD model is specified as:

$$Y_{it} = \alpha_i + \delta_t + \gamma D_{it} + \mathbf{X}\boldsymbol{\beta} + v_{it} \qquad \forall v_{it} = \sum_{j=1}^{p} \rho_j v_{it-j} + u_{it} \tag{11}$$

Here, α_i is a city-specific effect, δ_t is a time fixed effect, D_{it} is the status of intervention of unit i in period t, and $\mathbf{X}_{\mathrm{it}}$ is a vector of observed covariates. Unobserved components are summarized as v_{it}, which follows an AR(p) process because of serial correlations. Another reason to include lagged terms is to capture dynamic changes in outcomes over periods.

Figure 5 plots the trends of averaged mortality rates for the treated group and the control group. They were at similar levels around 1900. In the 1900s, the mortality rate of the treated group grew at a faster rate than that of the control group. Both groups reached their peaks in 1910. Afterward, both trajectories began to decline. Compared with control units, the treated units experienced faster decline in the 1910s to 1930. The result was mortality rate in the treated group were lower the controls in

the late 1920s. Obviously, the commend trend assumption for DD specification is not satisfied, since the two trends were not parallel before the intervention (Hastings 2004; Wolfers 2006). So SCM is preferred, since it requires no common trend assumption.

Table 6 reports outcomes four DD specifications. Effects of the 1916 intervention were in (1) and (2) are the same -0.7905, with only slight differences in their standard errors. Model (1) uses regular standard errors and model (2) uses clustered standard error. The coefficient of model (3) is -0.6865, since time fixed effects were not included. When I aggregate all observations into two periods, pre-1916 and post-1916, the estimate is -1.201, which is not significant at any conventional statistical level.

Unlike the MTSCM results in Table 4, DD models provide significant treatment effects using annual data. With the averaged outcomes, MTSCM suggested significant effects in only four years, 1924, 1926, 1927 and 1931. When aggregated data are used, DD estimation is no longer significant. As discussed, DD measures averaged gap between the treatment group and the control group before and after the intervention. But SCM focuses on the differences between the two groups after intervention, minimizing their discrepancy before intervention. For empirical SCM users, obviously, SCM is more useful to illustrate differences between treated units and their counterfactuals. DD results only tell us the averaged outcomes across periods and units. But SCM reveals differences across periods and treated units. For policy interpretation, the major benefit of SCM lies in presenting unit-specific treatment effects. For example, we can observe substantial and stable treatment effects in Chicago and Cleveland after 1916. However, such effects were not quite obvious in

Indianapolis and Milwaukee. Thus, the results explain why my parametric and non-parametric test statistics were not significant at any conventional level.

Figure 2- 5 Averaged Mortality Rate of Treated Units and Control Units

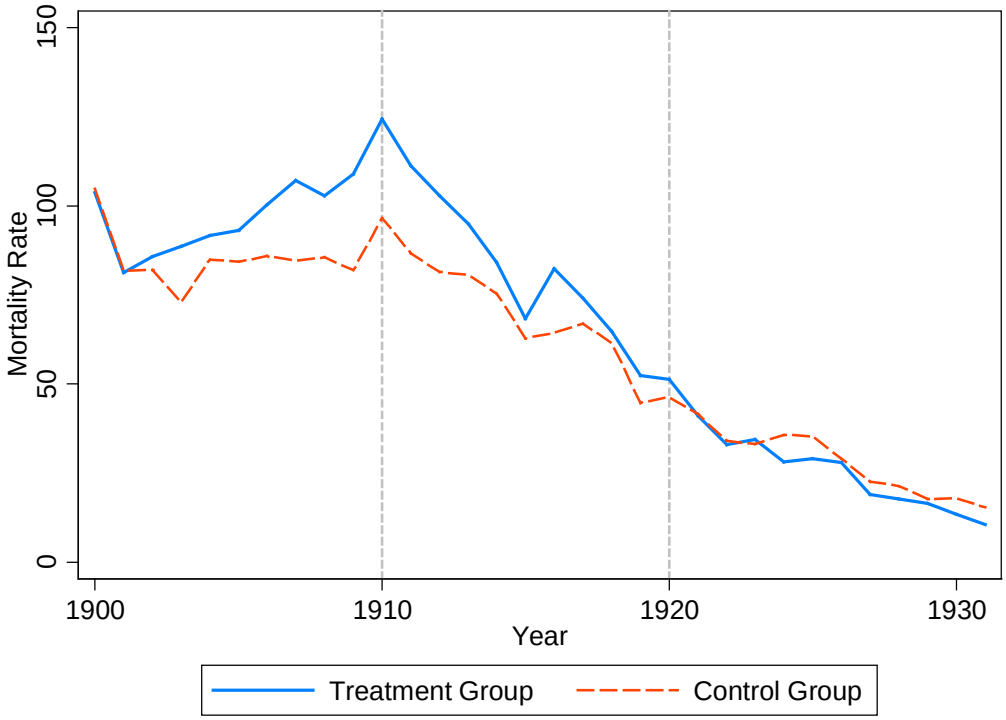


Table 2- 6 Difference-in-Difference Estimation of Pasteurization Ordinances 1916

Specifications	(1)	(2)	(3)	(4)
Variables				
Ordinance	-0.7905**	-0.7905**	-0.6865**	-1.2010
	(0.3690)	(0.3498)	(0.2780)	(1.1856)
Ln(Population)	0.3154	0.3154	0.0226	2.9164*
	(0.4960)	(0.5473)	(0.5993)	(1.6636)
Female Share	13.0429**	13.0429**	11.4980*	29.4723
	(5.9676)	(6.1768)	(6.3747)	(20.7750)
White Share	-8.5208***	-8.5208**	-13.3401***	-10.2539
	(2.8290)	(3.5937)	(2.9154)	(10.8071)
Average Income	0.0002	0.0002	-0.0036***	0.0008
	(0.0008)	(0.0010)	(0.0005)	(0.0038)
Mortality Rate -1	0.2833***	0.2833***	0.2973***	----
	(0.0246)	(0.0293)	(0.0293)	
Mortality Rate -2	0.2101***	0.2101***	0.2019***	----
	(0.0251)	(0.0345)	(0.0334)	
Mortality Rate -3	0.0635**	0.0635**	0.0706**	----
	(0.0248)	(0.0281)	(0.0269)	
Mortality Rate -4	0.1009***	0.1009***	0.1050***	----
	(0.0231)	(0.0335)	(0.0304)	
Observations	1,624	1,624	1,624	116
R-squared	0.6863	0.6863	0.6652	0.8071
Unit Fixed Effect	Yes	Yes	Yes	Yes
Time Fixed Effect	Yes	Yes	No	Yes

Notes: Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$
1. DD with regular standard errors
2. DD with clustered stand errors
3. DD with regular standard errors but no time fixed effects
4. DD with averaged outcomes before and after 1916

6. CONCLUSIONS

Focusing on public health, this study examines the causal effect of mandatory city milk pasteurization ordinances in the United States. As a remarkable food safety innovation, pasteurization was believed to be a critical factor in fighting epidemics in modern cities, with historical evidence showing its contribution to the decline of child diarrhea mortality in the early 1900s. However, pasteurization was also controversial because of competing interests between farmers, milk consumers and city health authorities. One key in this debate is to clarify the role of pasteurization with a persuasive causal health effect estimation and inference. However, such efforts have been rare in prior studies. This study aims to fill the void.

More than the causal health effects of pasteurization, there are two other focuses in this study. One is how to make valid inference with MTSCM. This study provides a case to show how to conduct inference for a small and non-normal sample in MTSCM applications. The other concern is the difference between DD and MTSCM estimators. SCM is regarded as a supplement of popular DD models, but they are based on different principles. A subtle line between them, DD can be taken as a special class of SCM when the counterfactual is constructed by equally weighted control units. SCM and MTSCM, however, use optimized weights.

Using MTSCM, this study measures causal health effects of pasteurization ordinances by combining unit specific and cross-unit evidence. In my sample, there are six cities that adopted ordinances in 1916 and 52 cities unexposed to similar interventions. Following a standard SCM algorithm, the intervention effect in each treated city is measured as the difference between the real and synthetic trajectories.

At the individual level, the results indicate Chicago and Cleveland had stable and substantial treatment effects after 1916. Some effects are also observed in Richmond and San Francisco, but none are noticeable in Milwaukee and Indianapolis. For inference, I choose two approaches. One is an "over-simplified" sample means comparison. In each post-intervention year, the averaged treatment effect (real-synthetic gap) is compared with the control group. Applying SW tests, I choose eight years in which control group outcomes are normal. The test statistics suggest treatment effects are significant in only four years. Then, I switch to non-parametric rank-sum tests which allow non-normal distribution and unpaired units. The results suggest that real-synthetic gaps in both groups are not statistically significant in every post-intervention year. For Robustness, I adopt a third approach which makes permutation based power test with repeated random sampling. The results are consistent with the rank-sum tests.

Using yearly data, DD estimations suggest treatment effects were noticeable and significant. For valid standard errors, I aggregate the panel into two periods, e.g. pre-1916 and post-1916. Using aggregated data, treatment effects are not significant. The comparison between MTSCM and DD results indicate researchers should be careful to interpret DD results in practice. Regarding DD's two sample means comparison, the major benefit of SCM application lies in presenting unit-specific treatment effects. One implication from SCM and MTSCM is unobserved heterogeneity could alternate estimations over periods. However, DD cannot reflect such time variant unobserved dependences. A substantial and significant DD

coefficient may be less useful to reveal the variations across units. That is why SCM was suggested as a supplement for DD.

Overall, this study suggests pasteurization was an important measure to save children from milk diseases in some cities but not consistently in all of them. Its potential health benefits could be still large, especially in cities with very low extent of pasteurization. For empirical SCM users, one key to extend this method for multiple units with treatment is to make valid inferences. Also, this study suggests regression based DD models could lead to different estimations as SCM does. Results suggest SCM reveals more information, *e.g*. unit-varying and time-varying treatment effect. This point is particularly meaningful for proper policy implications.

REFERENCES

Abadie, A., A. Diamond, and J. Hainmueller. “Synthetic Control Method for Comparative Case Study: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105, no. 490 (2010): 493–505.

Abadie, A., A. Diamond, and J. Hainmueller. “Comparative Politics and the Synthetic Control Method.” *American Journal of Political Science* (Article first published online: 23 APR 2014).

Abadie, A., and J. Gardeazabal. “The Economic Costs of Conflict: A Case Study of the Basque Country.” *American Economic Review* 93, no.1 (2003): 112–32.

Abadie, A., “Using Synthetic Controls to Evaluate an International Strategic Positioning Program in Uruguay” [Online], Available at: http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=37312864

Alvord, H. E. and R. A. Pearson, The milk supply of two hundred cities and towns, U.S. Department of Agriculture, Bureau of Animal Industry, Bulletin, 46, Washington, DC: 1903

Ayers, S. H . 1922. The present status of the pasteurization of milk, Bulletin No. 342. United States Department of Agriculture

Ayers, S. H . 1926. The present status of the pasteurization of milk, Bulletin No. 342. revised edition. United States Department of Agriculture

Beaver, M. W. “Population, infant mortality and milk.” Population Studies 27, no. 2 (1973): 243-254.

Bertrand, M., E. Duflo, and S. Mullainathan. “How Much Should We Trust Differences-in-Differences Estimates?.” *The Quarterly Journal of Economics* 119, no. 1 (2004): 249-275.

Billmeier, Andreas, and Tommaso Nannicini. “Assessing economic liberalization episodes: A synthetic control approach.” *Review of Economics and Statistics* 95, no. 3 (2013): 983-1001.

Block, D. R. “Public health, cooperatives, local regulation, and the development of modern milk policy: the Chicago milkshed, 1900–1940.” *Journal of Historical Geography* 35, no. 1 (2009): 128-153.

Block, D. “Purity, economy, and social welfare in the progressive era pure milk movement.” Journal for the Study of Food and Society 3, no. 1 (1999): 20-27.

Cameron, A. Colin, and Pravin K. Trivedi. Microeconometrics: methods and applications. Cambridge university press, 2005.

Cutler, D., and G. Miller. “The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.” *Demography* 42, no.1 (2005): 1–22.

Czaplicki, A . “Pure Milk Is Better Than Purified Milk Pasteurization and Milk Purity in Chicago, 1908-1916”, *Social Science History* 31: (2007) 411-433

Dube, A., and B. Zipperer. “Pooled Synthetic Control Estimates for Recurring Treatments: An Application to Minimum Wage Case Studies.” Unpublished paper (2013)

Fares, M. and E. Rouviere. “The implementation mechanisms of voluntary food safety systems.” Food Policy 35, no. 5 (2010): 412-418.

Frank, L. C. and Moss, F. J . 1931. The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931. US Public Health Service

Gobillon, L., and T. Magnac. “Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls.” (2013) Unpublished Paper.

Hall, C. W. and Trout G M . 1968. Milk pasteurization. Westport, Conn., Co

Hastings, Justine S. “Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California.” *American Economic Review* 94.1 (2004): 317-328.

Henson, S., and J. Caswell. “Food safety regulation: an overview of contemporary issues.” *Food policy* 24, no. 6 (1999): 589-603.

Hoover, Kevin D. “Causality in economics and econometrics.” The new Palgrave dictionary of economics 2 (2008)

Imbens, G. and Wooldridge J., Lecture Notes 10, National Bureau Economic Research (NBER) Summer Institute Econometric Lectures,

Kelly, E, and Clement C E . 1931. Market milk. J. Wiley & Sons, Inc: New York; Chapman & Hall Limited: London

Lee, K-S. “Infant mortality decline in the late 19th and early 20th centuries: the role of market milk.” Perspectives in biology and medicine 50, no. 4 (2007): 585-602.

Levitt J. W., The Healthiest City: Milwaukee and the Politics of Health Reform, 1996, p. 187

Mann, H. B., and D. R. Whitney. “On a test of whether one of two random variables is stochastically larger than the other.” *The Annals of Mathematical Statistics* (1947): 50-60.

Meckel R. A., Save the Babies: American public health reform and the prevention of infant mortality, 1850-1929, Johns Hopkins University Press, 1990

Olmstead, A. L., and P. W. Rhode. “An impossible undertaking: the eradication of bovine tuberculosis in the United States.” *The Journal of Economic History* 64, no. 03 (2004): 734-772.

Rodrik, D. One economics, many recipes: globalization, institutions, and economic growth. Princeton University Press, 2008.

Segerson, K. “Mandatory versus voluntary approaches to food safety.” Agribusiness 15, no. 1 (1999): 53-70.

Selitzer R . 1976. The dairy industry in America. Magazines for Industry: New York

Shapiro, S. S., and M. B. Wilk. “An analysis of variance test for normality (complete samples).” *Biometrika* (1965): 591-611.

Vögele, Jörg, and Wolfgang Woelk. “Public Health and the Development of Infant Mortality in Germany, 1875–1930.” *The History of the Family* 7, no.4 (2002): 585–99.

Winslow C-E A . 1952. Man and epidemics, Princeton University Press. Princeton: New Jersey

Wolf, J. H. “Low Breastfeeding Rates and Public Health in the United States.” *American Journal of Public Health* 93, no. 12 (2003): 2000–10.

Wolfers, Justin. “Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results.” *American Economic Review* 96.5 (2006): 1802-1820

CHAPTER III.

ROBUSTNESS TESTS FOR CAUSAL EFFCTS OBTAINED FROM SYNTHETIC CONTROL ESTIMATIONS: A CROSS-SECTIONAL TIME SERIES MODLE

Huiqiang Wang

Department of Environmental and Natural Resource Economics,

University of Rhode Island, Kingston, RI, USA 02881

July 2015

ABSTRACT

This paper aims to provide a robustness test for major conclusions obtained from prior chapters, *e.g*. the effect of Chicago's 1916 milk pasteurization ordinances. Using the synthetic control methods (SCM), I found a significant treatment effect. To verify SCM results, I use a linear regression based cross-sectional time series model (CTM) to re-estimate this intervention. CTM results confirm major findings in my prior SCM studies. In addition, I use the 1989 California cigarette sales tax as an "out-of-sample" robustness check for CTM. Again, CTM results are similarly significant as SCM. This study provides some evidence CTM could be an option for validating SCM results in practice.

Keywords: Pasteurization Ordinance, Synthetic Control Methods, Cross-sectional Time Series Model, Robustness Tests

1. INTRODUCTION

Historical evidence sheds light on the contribution of cow milk to human health and nutrition, particularly in the transition of early childhood mortality in the early 20th century (Beaver 1973). In the mid-1800s, the milk supply in many cities in the United States was riddled with adulterations and biological contaminations. Lower quality milk was a major source of epidemics, for example diarrhea, tuberculosis, scarlet fever and sore throat (Seltzer 1976). After the introduction of pasteurization, milk finally became a safe source of nutrition instead of a health threat. Voluminous prior literatures suggested the critical role of pasteurization in all these changes, especially in populous cities (Meckel 1990; Vögele and Woelk 2002; Wolf 2003; Lee 2007). After initial voluntary commercial implementations, pasteurization was incorporated into regulations of health departments in some cities. Health professionals have since lauded these changes as key step in the influential public health campaign of the early 20th (Cutler and Miller 2005). However, mandatory pasteurization ordinances also caused opposition. Consumers worried about possible nutrient losses, physical and flavor changes of pasteurized milk, and long-term health impacts to children (Hall and Trout 1968). At the same time, the issue was even more complicated since interest conflicts occurred between dairy farmers and city health officials.

The health impacts of pasteurization have been a key issue. Researchers are curious to know whether pasteurization policies made a substantial difference to health outcomes. Prior studies do not give us a rigorous quantitative conclusion on the effects of pasteurization ordinances, so the previous two chapters serve to make a clear and consistent causal estimate of the health impact of these policy interventions.

The first two chapters discussed the causal effects of Chicago's 1916 pasteurization ordinance and similar interventions in five other cities in that year. Two empirical methodologies are synthetic control methods (SCM) and Difference-in-Difference (DD) models. For example, single unit DD comparison was made between Chicago and St. Paul in Chapter 1. Then SCM estimation was made by comparing Chicago with its synthetic version, constructed from 20 control units. In Chapter 2, I extended SCM to multiple treated units (MTSCM) and examined the effect of 1916 ordinances in Chicago, Cleveland, Indianapolis, Milwaukee, Richmond and San Francisco. In addition, I made DD estimations and compare them to MTSCM results. Some major findings in Chapter 2 are as follows. First, SCM results show health effects of Chicago's 1916 ordinance was obvious and significant. Second, unlike the case of single treated unit, MTSCM results differed across treated units. Significant effects were found in some cities, but not all. Third, statistical inference in MTSCM needs more attention. Considering sample sizes and distribution of outcomes, nonparametric methods are preferred in this study. Last but not least, SCM and MTSCM can be supplements for DD in empirical applications (Abadie, Diamond and Hainemuller, ADH 2010), since they reveal more information than DD.

As an innovative econometric tool for comparative case studies, SCM applications have been increasing in recent years. SCM was proposed by Abadie and Gardeazabal (2003) and became mathematically formalized in ADH (2010, 2014). Being a bridge between quantitative and qualitative studies, SCM also connects unit-specific and cross-sectional evidence (Billmeier and Nannicini 2013). It can be applied to studies at both macro and micro levels. For macro-level studies, Abadie and

Gardeazabal (AG 2003) applied SCM to estimate economic costs of conflict in the Basque country using regions without terrorist conflicts in Spain. Billmeier and Nannicini (2013) used this method to investigate the impact of economic liberalization on real GDP per capita with a worldwide sample. Also, ADH (2014) explored economic costs of Germany 1991 re-unification by making a synthetic version with a small sample of OECD countries. More than applications to aggregate units with macroeconomic data, SCM was also used for micro-levels cases. Kiesel and Villas-Boas (2010) measured the effect of nutritional labels to consumers' choices in stores. Pooling multiple treated units, Dube and Zipperer (2013) studied the effect of recurring treatment on the minimal wage changes in 45 states in the US.

Overall, empirical benefits of SCM are four-fold. First, SCM is a good tool for analysis of aggregated entities, as it requires data at aggregated levels for estimation. Second, SCM provides users a variety of inferential toolkits and robustness diagnosis and validity tests. Third, with a non-parametric algorithm, SCM provides a systematic way to select control units. It generates explicit weights which are constrained as positive and summed to one. This unique feature allows SCM users to interpret the weight as the specific contribution of each control units. Finally, SCM provides user-friendly visualizations. Graphically, researchers can illustrate how treatment effects vary across periods.

However, SCM models are contextually restrictive in some applications. Particularly, they have two requirements that need to be satisfied (Abadie online). The first one is the "convex hull condition", namely characteristics of the treated unit should be comparable with units unexposed to the intervention. Second, SCM prefers

low volatility of outcomes. Volatile outcomes could blur small intervention effects and random shocks (Abadie online). Thus, for empirical users, it is desirable to have alternative specifications that can make robustness check to SCM results. This chapter is intended to investigate whether results in prior chapters can be supported by alternative econometric methods.

Using cross-sectional time series model (CTM), Hsiao, Ching and Wan (HCW 2012), introduced a counterfactual building algorithm for comparative case studies. As we know, early comparative case studies are based on DD specifications from Card (1990) and Card and Kruger (1994). These models stressed the "similar trends" assumption. Later, new attentions are given to selecting proper control units, for example the SCM (AG 2003; ADH 2010; 2014). The key of SCM is to build a counterfactual with the optimally weighted cross-sectional units. Motivated by a similar principle but different focus, HCW (2012) introduced CTM method, which is based on linear regressions. One major advantage of CTM, it is computationally easier than SCM, since it requires only outcomes for regression. A second benefits, it provide an approach to avoid over-fitting in other linear specification. In sum, SCM and CTM share some common features. On one hand, a linear combination of optimally selected control units is better than any single unit as a counterfactual. On the other hand, an optimal subset is more reliable than the model which includes all comparison units.[54]

This paper is organized as follows. Part 2 introduces the CTM model of HCW (2012). It covers CTM set-up and some major assumptions. Using Chicago as a case

[54] In the case of a small comparison period or large number of control units, we need some procedure to reduce the number of control units to meet the dimensional requirement.

study, Part 3 compares performances of SCM and CTM in an empirical context. The result would provide a robustness test to prior SCM conclusions. Next, Part 4 will make another CTM application to the cigarette sale tax of ADH (2010) as an "out-of-sample" check to the efficiency of this method. Finally, concluding remarks are wrapped in Part 5.

2. CROSS-SECTIONAL TIME SERIES MODEL

HCW (2012) proposed an alternative for SCM specifications, using outcome variables only. Here it is referred as the CTM. Suppose outcomes of treated unit *i* and units *j* unexposed are Y_{it} and Y_{jt}, intervention D_{it} occurred at *T+1*, and X_{it} is a vector of control variables (common factors) which varies over *t*.[55] If specific assumptions are satisfied, CTM can predict $\hat{Y}_{it}^{0}$ using pre-intervention outcomes of control units Y_{jt} only, in which the information of X_{it} has been embedded.[56] Next, an unobserved counterfactual is estimated by a linear combination of its control units. More details about these empirical assumptions are in Bai and Ng (2002) and HCW (2012). These assumptions includes,

Assumption 1: $E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$; $E(\varepsilon_{it}|D_{js}) = 0$ for $i \neq j$ and $s \neq t$.

Assumption 2: β is full rank and $\|\beta_i\| = c < \infty$ for all *i*.

Assumption 3: $E(\varepsilon_{it}|X_{it}) = 0$

Assumption 4: $E(\varepsilon_{it}|D_{js}) = 0$ for $i \neq j$ and $s \neq t$.

Remark 1: We assume ε_{it} is a white noise process and uncorrelated with common factors and treatment variable. ε_{it} are uncorrelated across units. The effects of common factors can vary across units, *e.g.* allowing $\beta_i \neq \beta_j$ (HCW 2012).

Remark 2: HCW made no assumption on the time series properties of Y_t and X_t. As for time series properties, X_t can be either stationary or non-stationary. For Y_t, their

[55] The work of Hsiao *et al*. (2012) follows the fashion of linear regression and can be taken as an exception of Abadie *et al*. (2014)'s comments to regression methods in practice.
[56] The assumptions are in Hsiao *et al*. (2012) Assumption (1) to (5). See proposition 1 in Appendix.

model implies the outcome[57] follows an autoregressive moving average (ARMA) model. X_t, Y_t can be stationary or non-stationary. Similarly, SCM has no stationarity requirement for the data. Abadie (online) pointed out SCM should be less appropriate if the outcomes are very volatile. In many recent SCM applications, we find the outcome variables usually have unit-root. The treated unit is non-stationary.

$$Y_{it}^{0} = w' Y_{jt} + r_{git} + v_{it} \text{ for } t \in [1, \cdots, T] \tag{1}$$

Optimized weights (w_j) are obtained by minimizing discrepancy between the actual Y_{it} and e $\hat{Y}_{it}^{0}$ which is a linear combination of y_{jt}, as (6)

$$(w_j) = \arg\min_{w \in \mathbb{R}^T} \frac{1}{T} [(y_{it} - w' Y_{jt})' A (y_{it} - w' Y_{jt})] \tag{2}$$

Remark 3: In CTM, cross-sectional control units selection are empirically based on R^2or likelihood. SCM algorithm of ADH (2010) is maximum likelihood estimation which constrains weights to be positive and to sum to 1. The CTM, on the other hand, applied the least square regression to the set of control units selected by post intervention mean square prediction error (PMSE), Akaike information criteria (AIC), Bayesian Information Criteria (BIC) and corrected AIC (AICC). Weights of control units in their model have fewer restrictions, allowing negative weights and the sum does not equals one.

The dimensional issue could also be a concern if the number of comparison units exceeds the number of pre-intervention observations. In addition, HCW (2012) point out there will be no limit for the number of cross-sectional units (*N*), if we have a large

[57] HCW (2012) P.712 Assumption 7, the authors supposed the treatment effect ($\hat{\Delta}_{it}$) follows the autoregressive (AR) process. As we can see the treatment effect is in fact a linear combination of the outcomes of the treated unit and its control group. So we see outcome variables also follow AR process.

pre-intervention sample ($T \longrightarrow \infty$). But when N/T is finite, Hsiao *et al*. recommend using a subset (p) of N, which can provide optimal $\hat{Y}_{it}^{0}$. The choice of p involves two steps. First, units j ($j \in N$) which are the best predictors of $\hat{Y}_{it}^{0}$ are selected via likelihood or R^2. Second, using Akaike information criteria (AIC), corrected AIC or similar criteria, users choose the best group of predictors. Their simulations suggest that the chosen set is better than the model using all control units in prediction.[58]

ADH (2014) noticed the connection between SCM and regression based methods. For SCM, its major benefit is to provide an explicit algorithm in selecting control units. Ordinary least squares (OLS) regression could have lower pre-intervention error, but may have the risk of over-fitting. Unlike OLS, CTM uses a selected subset instead of all units to construct the counterfactual. [59] Similar to OLS, there is no guarantee that CTM weights are in the range of zero to one and summed to one like SCM. In this study, I modify HCW's specification into a three step process.[60]

(A) Suppose I have a finite pre-intervention period, and need to select a subset of control units. The nature of SCM and CTM is to use a proper comparison group to trace the real trend before intervention. So it is ideal to use those units which can approximate the treated unit as close as possible.

[58] More discussions are in HCW (2012) and Hsiao and Wan (2014)
[59] As discussed, if there is no dimensional issue.
[60] One reason to modify HCW model is the dimensional issue. Based on quarterly data, dimension was not a serious concern in HCW (2012). The number of observations is much larger than the number of units in the control group. For cases in this study, Chicago and California, annual pre-intervention data are small. So I need a pre-step to select proper control units.

$$
\begin{aligned}
&Y_{it}^{0} = A_j Y_{jt} + u_{jt} \text{ for } j = 1, \cdots, N \text{ and } t \in [1, \cdots, T] \\
&Y_{it}^{0} = A_1 Y_{1t} + u_{1t} \quad \Rightarrow AIC(1),\ R^2(1) \\
&\vdots \\
&Y_{it}^{0} = A_N Y_{Nt} + u_{Nt} \ \Rightarrow AIC(N), R^2(N)
\end{aligned} \tag{3}
$$

We then select the units (k) with the smallest AIC values (or the highest R^2), which can better approximate the unit exposed to intervention. These units are ranked by their AIC values from the highest to the smallest, as AIC(1), ⋯, AIC(k).

(B) The selected k units are then used as predictors for Y_{it}^{0} with k specifications.

$$
\begin{aligned}
&Y_{it}^{0} = A_1 Y_{jt} + u_{jt} && \Rightarrow AIC(1) \\
&Y_{it}^{0} = A_1 Y_{1t} + A_2 Y_{2t} + u_{2t} && \Rightarrow AIC(2) \\
&Y_{it}^{0} = A_1 Y_{1t} + A_2 Y_{2t} + A_3 Y_{3t} + u_{3t} && \Rightarrow AIC(3) \\
&\vdots \\
&Y_{it}^{0} = A_1 Y_{1t} + A_2 Y_{2t} + A_3 Y_{3t} + \cdots + A_k Y_{kt} + u_{kt} && \Rightarrow AIC(k)
\end{aligned} \tag{4}
$$

In other words, predictors (control units) are one-by-one added to the regression model regarding their closeness to the treated unit before the intervention. Similar as HCW, AIC values choose an optimal group of predictors.[61] Their Simulations also suggest a subset (instead of all control units) has a lower AIC when the number of pre-intervention period is definite.

(C) Using the selected units, OLS is used to generate the weights of each control units.

$$
\begin{aligned}
&(\tilde{w}_k) = \arg\min_{w \in \mathbb{R}^T} \frac{1}{T} [(Y_{it} - w_k' Y_{kt})' A (Y_{it} - w_k' Y_{kt})] \\
&\text{where } \ S_k \subseteq S_j (k \leq N)
\end{aligned} \tag{5}
$$

[61] Like other model selection criteria, irrelevant regressors would inflate R^2 but decrease AIC values.

The obtained weights are then applied to construct the counterfactual and estimate the effect of intervention as (6).

$$\hat{\hat{\Delta}}_{it} = Y_{it} - \hat{Y}_{it} = Y_{it} - \sum_{i=T}^{T+P} Y_{kt}\tilde{w}_k \text{ for } t = T+1, \cdots, T+P \qquad (6)$$

The quality of fitting can be measured by the root of mean squared prediction error (RMSPE) as (7).

$$RMSPE = \left[\frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \sum_{j=1}^{J} y_{jt} w_j \right)^2 \right]^{1/2} \text{ for } t = [1, \cdots, T] \qquad (7)$$

3. CHICAGO 1916 ORDINANCE: A REVISIT

To verify SCM estimation in prior chapters, this paper will make a CTM robustness check using Chicago's 1916 pasteurization ordinance. As mentioned, one rationale to use outcome variables as predictors in HCW (2012) is we assume information of other covariates has been embedded into outcomes. Thus, this study will compare CTM results with SCM models when non-outcome covariates exist.

Table 1 illustrates predictors used to construct synthetic versions for Chicago's 1916 policy intervention. For simplicity, here I choose one set of covariates without formal cross-sets comparisons.[62] In this set, I include female share, white share, share of population under 5 year old, average income and four pre-intervention outcomes in 1900, 1905, 1910 and 1915. Non-outcome predictors are averaged between census years 1900 and 1910. Outcome of interests is the same mortality rates as I used in prior two chapters. [63] For SCM estimations, there are two choices, e.g. using a nested optimization (SCM 1) and regular algorithm (SCM 2).

Real and synthetic values of selected predictors are presented in Table 1. SCM 1 and SCM 2 values approximate non-outcome covariates and outcome predictors quite well. Differences between non-outcome covariates are quite small. For outcome predictors, they are slightly different from the real values. Weights generated from the two SCM models are in Table 2. We can see the weights generated are not quite consistent. For example, the weight of New Haven was zero in SCM 1 and 0.321 in SCM 2. San Diego, on the other hand, was 0.286 in SCM 1 but zero in SCM 2. Other cities are also differently weighted. Figure 1 and Figure 2 plot the two synthetic

[62] More details are in Dube and Zipper (2013). The authors set up a five step process to choose a best set of predictors.
[63] Data sources are the same as in Chapter 1 and Chapter 2. Details can be found in these two papers.

trajectories. It is clear that the fitting of SCM 1 is better than SCM 2. A RMSPE comparison between them will be discussed later.

Table 3- 1 Predictor Values of SCM Models: Chicago 1916

Predictors	Chicago	SCM 1	SCM 2
Female share	0.49	0.49	0.49
White share	0.98	0.93	0.92
Percent aged < 5 year	0.11	0.09	0.10
Per capita wage	550.53	549.95	545.81
Mortality rate 1900	9.37	9.65	10.31
Mortality rate 1905	10.82	9.01	10.26
Mortality rate 1910	15.69	15.69	15.21
Mortality rate 1915	11.02	12.49	11.34

Note: SCM 1 – synthetic control methods with nested algorithm; SCM – regular algorithm

Table 3- 2 SCM Weights of Control Units: Chicago 1916 Intervention

City	SCM 1	SCM 2	City	SCM 1	SCM 2
Duluth	0.034	0.005	Omaha	0	0
Evansville	0	0	Portland	0	0
Hartford	0	0	Providence	0.003	0
Jacksonville	0	0	San Antonio	0.334	0.339
Kansas city	0	0	San Diego	0.286	0
L. Angeles	0	0	Seattle	0	0
Memphis	0	0	St. Paul	0.207	0
Nashville	0	0	Tacoma	0	0.168
New Haven	0	0.321	Utica	0.138	0
N. Orleans	0	0	Wichita	0	0.168

As discussed, CTM uses a unit-selection process to avoid over-fitting. Following (3) and (4), three control units are selected using AIC and BIC. They are San Antonio, New Orleans and Utica (Table 3). Their weights are generated from ordinary least square (OLS) regression with pre-intervention outcomes of Chicago. CTM trajectory is depicted in Figure 3. A formal fitting comparison of SCM 1, SCM 2 and CTM is available in Table 4. Results indicate SCM 1 and CTM have similar pre-intervention RMSPE, while the one of CTM is slightly lower. Both fittings are better than SCM 2. For post-intervention fittings, the three specifications have similar RMSPE.

Table 3- 3 Weights of CTM: Chicago 1916 Intervention

	Coefficient	SD	*t*-stat.	*p*-value
San Antonio	0.252	0.049	5.190	0.000
New Orleans	0.256	0.090	2.840	0.014
Utica	0.303	0.086	3.530	0.004

For statistical inferences, I use bootstrap methods to construct standard errors for SCM 1, SCM 2 and CTM. Their results are in Table 5. Coefficients measure real-synthetic gaps in 1930. We also have bootstrapped standard errors and *p*-values. Using different control units, coefficients of interests are different in the three models. However, all of them indicate a significant treatment effect in 1930. In sum, CTM results confirmed SCM conclusions, e.g. the 1916 intervention effect was significant. In the next part, I will use the case study of California's 1989 cigarette tax for an "out of sample" test for robustness of CTM estimation in practice.

Table 3- 4 RSMPE in Different Specifications: Chicago 1916 Intervention

RMSPE	SCM 1	SCM 2	CTM
Pre-intervention	1.0928	1.9060	0.9576
Post-intervention	2.6003	2.5235	2.6729
Post/Pre-ratio	2.3795	1.3240	2.7911

Figure 3- 1 Real-Synthetic Trajectories of Chicago Intervention 1916 SCM 1

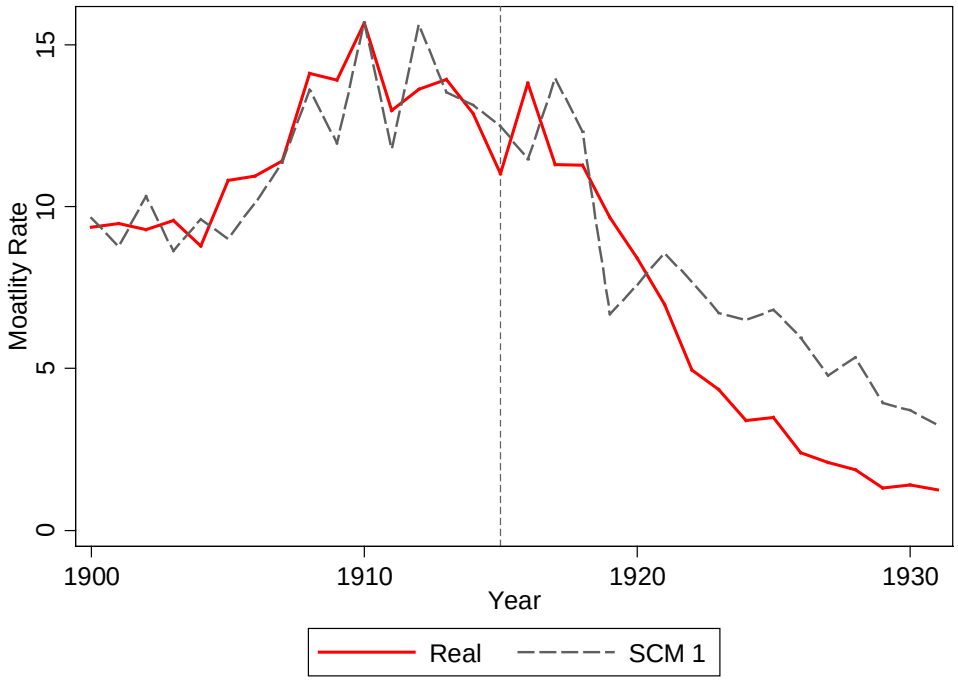


Figure 3- 2 Real-Synthetic Trajectories of Chicago Intervention 1916 SCM 2

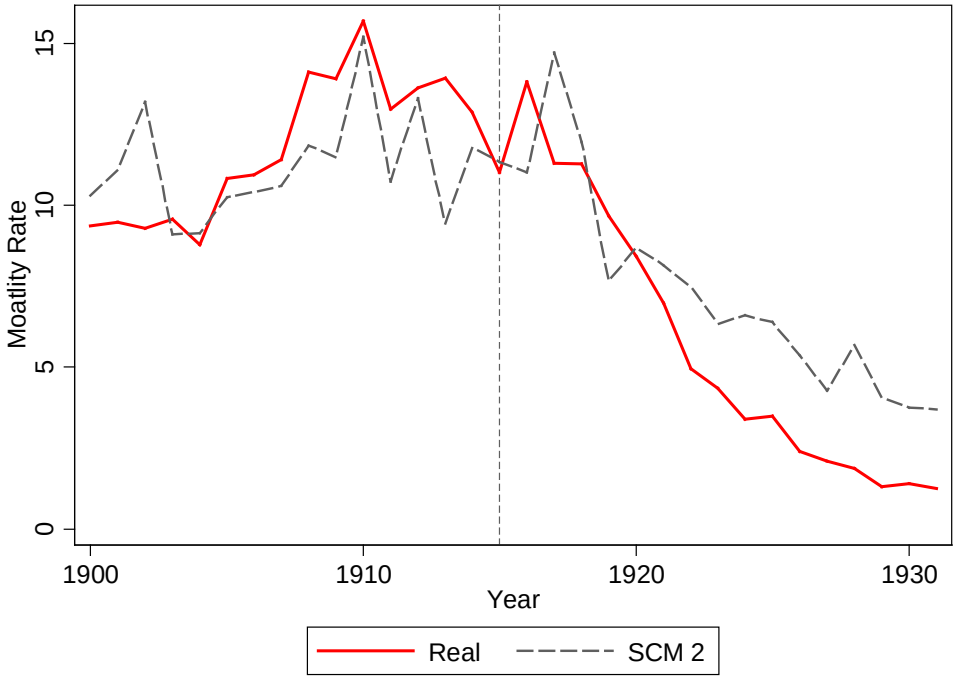


Figure 3- 3 Real-Synthetic Trajectories of Chicago Intervention 1916 CTM

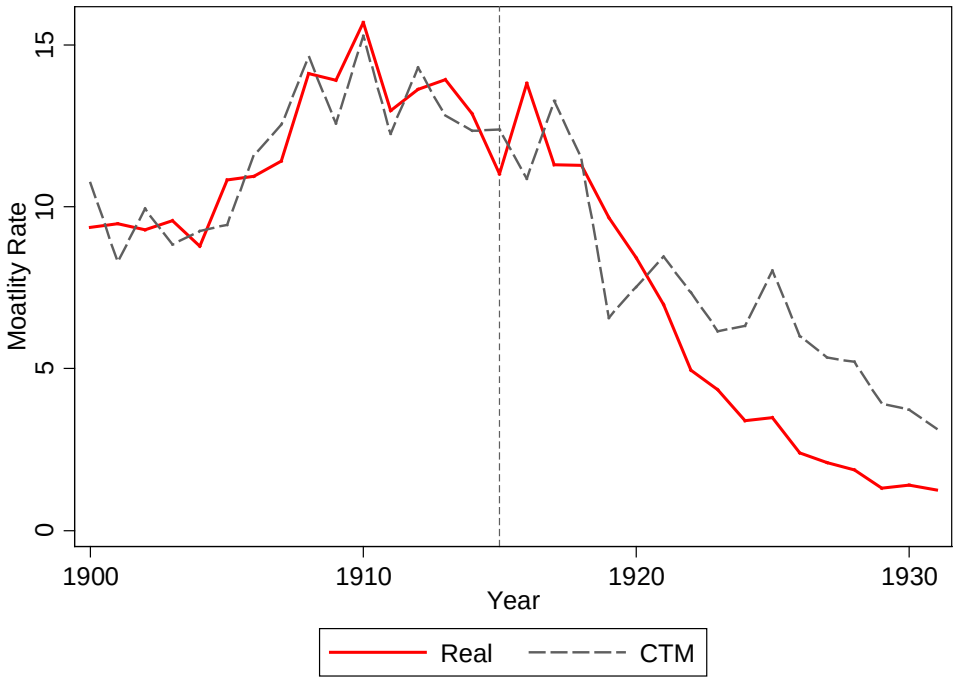


Table 3- 5 Bootstrap Re-sampling of Treatment Effect: Chicago 1916

	Repetitions	Coefficient	Bootstrap S.D.	Z	*p*>Z
SCM 1	N=50	-1.599	0.708	-2.260	0.024
	N =100	-1.599	0.602	-2.660	0.008
	N =500	-1.599	0.768	-2.080	0.037
	N =1000	-1.599	0.640	-2.500	0.012
SCM 2	N=50	-2.599	0.665	-3.910	0.000
	N =100	-2.599	0.734	-3.540	0.000
	N =500	-2.599	0.698	-3.720	0.000
	N =1000	-2.599	0.598	-4.350	0.000
CTM	N=50	-2.329	0.389	-5.990	0.000
	N =100	-2.329	0.291	-8.000	0.000
	N =500	-2.329	0.325	-7.180	0.000
	N =1000	-2.329	0.317	-7.360	0.000

Note: SCM 1 generates 6 control units with non-zero weights; SCM 2 generates 5 control units with non-zero weights; CTM generates 3 control units with non-zero weights

4. CALIFORNIA CIGARETTE SALES TAX

For robustness, I will make another empirical application to compare CTM and SCM estimations. This case is the 1989 California Cigarette Sales Tax. ADH (2010) used this policy intervention in their formal SCM analysis. Details of this legislation can be found in ADH (2010). As anti-tobacco legislation, California initiated Proposition 99 in 1988 to increase California's cigarette excise tax by 25% per pack. The increased revenues will be used for anti-tobacco projects in the state. Using yearly state level panel data 1970-2000 and SCM, ADH measured the impact this policy intervention to per capita cigarette sales in California after 1988. Their predictors include GDP per capita (log values), percent aged 15-24, cigarette retail prices, beer consumption per capita, and outcomes of interest in 1975, 1980 and 1988. California is the unit exposed to intervention. Control units include 38 states without similar policy interventions in the observation period. SCM results indicated California's real cigarette sales were much lower than the synthetic version after implementation of this proposal. And the real-synthetic gap in California was significantly larger than its control units which are randomly generated from placebo studies.

Similar to the procedure in the prior section, I will apply CTM to re-estimate SCM results in ADH (2010). If their results are close, this case study would provide one more support for CTM as an alternative of SCM. Also, it helps to confirm robustness of the Chicago study. There are two SCM specifications as above. The first one is the nested SCM models (SCM 1). The second uses the same predictors but regular SCM algorithm (SCM 2). Predictor values of SCM 1 and SCM 2 are reported in Table 6. Weights generated are presented in Table 7. As Table 6 suggests, SCM 1 is

very close to the results reported in ADH (2010) with only decimal differences. SCM 2 and SCM 1 are similar, with only slight differences in non-outcome predictors. SCM 1 generated five units with non-zero weights, while SCM 2 had four states which have positive weights. SCM 1 states are Colorado, Connecticut, Montana, Nevada and Utah. In SCM 2, Montana has zero weight (Table 7). CTM units are Colorado, Illinois, New Hampshire and Nevada (Table 8). Compared with SCM models, CTM has a smaller pre-intervention RMSPE, which implies better fitting of real and synthetic trajectories (Table 9). In addition, post-intervention RMSPE of CTM is the smallest among the three models.

Table 3- 6 Predictor Values of SCM Models: California 1989

Predictors	California	SCM 1	SCM 2
Ln(GDP per capita)	10.08	9.86	9.90
Percent aged 15-24	89.42	89.41	89.00
Retail price	0.17	0.17	0.18
Beer per capita	24.28	24.22	23.26
Cigarette sales 1975	127.10	127.14	126.39
Cigarette sales 1980	120.20	120.59	120.72
Cigarette sales 1988	90.10	91.76	92.09

Note: SCM 1 – synthetic control methods with nested algorithm; SCM – regular algorithm

Real and synthetic trajectories of these three models are plotted in Figure 4 to Figure 6. Overall, the three synthetic trajectories had a similar trend. Noticeable real-synthetic gaps can be observed in all three models. In comparison to SCM 1 and SCM 2, CTM has a better fitting as Table 9 illustrates. For inference, Table 10 presents bootstrap standard errors of treatment effect in 2000. The treatment effects are all significant. So the results suggest CTM can be an alternative for SCM estimators.

Aside from the performance of CTM and SCM, the two studies highlighted one contextual requirement of SCM, the volatility of outcomes. SCM is preferable for outcomes with lower volatility, because highly volatile results might not be distinguishable from random errors. As Abadie (online) explained, "The nature of this exercise, which focuses on a single unit, indicates that small effects will be indistinguishable from random shocks to the outcome of the affected country, especially if the outcome variable of interest is highly volatile". Although no explicit time series property discussions were made in AG (2003) and ADH (2010, 2014), cigarette sales in California was not quite volatile. Table 11 shows the realization of California's cigarette sales before 1989 can be modeled as an AR (2) process with a linear time trend as below.

$$Y_{it} = 4314.45 + 1.473 Y_{it-1} - 0.577 Y_{it-2} - 2.123 \cdot t + e_{it} \quad \text{(California 1989)}$$

A Dickey-Fuller test suggests that pre-intervention Y_t is non-stationary. As a result, real and synthetic trends fit each other quite well before the intervention.

On the other hand, SCM fittings may be worse when the data generating process (DGP) switches to a volatile one. For example, DGP of Chicago's children mortality rates before 1916 can be specified as an AR (1) process, as Table 12.

$$Y_{it} = 11.271 + 0.777 Y_{it-1} + e_{it} \quad \text{(Chicago 1916)}$$ [64]

In the case of Chicago, the dependence of prior values is about 0.777, which explains its non-smooth trend before intervention. In sum, the two different DGPs in California and Chicago are corresponding to different real-synthetic fittings in Figures 1-3 and Figures 4-6.

[64] The coefficients of AR(1) terms are even lower in other specifications.

Table 3- 7 SCM Weights of Control Units: Chicago 1916 Intervention

State	SCM 1	SCM 2	State	SCM 1	SCM 2
Alabama	0	0	Nevada	0.234	0.217
Arkansas	0	0	New Hampshire	0	0
Colorado	0.164	0.356	New Mexico	0	0
Connecticut	0.069	0.083	N. Carolina	0	0
Delaware	0	0	N. Dakota	0	0
Georgia	0	0	Ohio	0	0
Idaho	0	0	Oklahoma	0	0
Illinois	0	0	Pennsylvania	0	0
Indiana	0	0	Rhode Island	0	0
Iowa	0	0	S. Carolina	0	0
Kansas	0	0	S. Dakota	0	0
Kentucky	0	0	Tennessee	0	0
Louisiana	0	0	Texas	0	0
Maine	0	0	Utah	0.334	0.344
Minnesota	0	0	Vermont	0	0
Mississippi	0	0	Virginia	0	0
Missouri	0	0	W. Virginia	0	0
Montana	0.199	0	Wisconsin	0	0
Nebraska	0	0	Wyoming	0	0

Table 3- 8 Weights of CTM: Chicago 1916 Intervention

	Coefficient	SD	*t*-stat.	*p*-value
Colorado	0.100	0.095	1.050	0.309
Illinois	0.359	0.067	5.370	0.000
New Hampshire	0.102	0.037	2.730	0.015
Nevada	0.186	0.043	4.290	0.001

Table 3- 9 RSMPE in Different Specifications: California 1989 Intervention

RMSPE	SCM 1	SCM 2	CTM
Pre-intervention	1.7563	2.0373	1.3455
Post-intervention	20.7285	19.5225	17.3082
Post/Pre-ratio	11.8022	9.5826	12.8636

Table 3- 10 Bootstrap Re-sampling of Treatment Effect: California 1989

	Repetitions	Coefficient	Bootstrap S.D.	Z	*p*>Z
SCM 1	N=50	-27.984	1.578	-17.730	0.000
	N =100	-27.984	1.178	-23.750	0.000
	N =500	-27.984	1.439	-19.450	0.000
	N =1000	-27.984	1.282	-21.830	0.000
SCM 2	N=50	-24.170	0.605	-39.930	0.000
	N =100	-24.170	0.644	-37.540	0.000
	N =500	-24.170	0.592	-40.850	0.000
	N =1000	-24.170	0.599	-40.370	0.000
CTM	N=50	-23.293	0.733	-31.780	0.000
	N =100	-23.293	0.717	-32.510	0.000
	N =500	-23.293	0.749	-31.100	0.000
	N =1000	-23.293	0.734	-31.730	0.000

Note: SCM 1 generates 5 control units with non-zero weights; SCM 2 generates 4 control units with non-zero weights; CTM generates 4 control units with non-zero weights

Figure 3- 4 Real-Synthetic Trajectories of California Intervention 1989 SCM 1

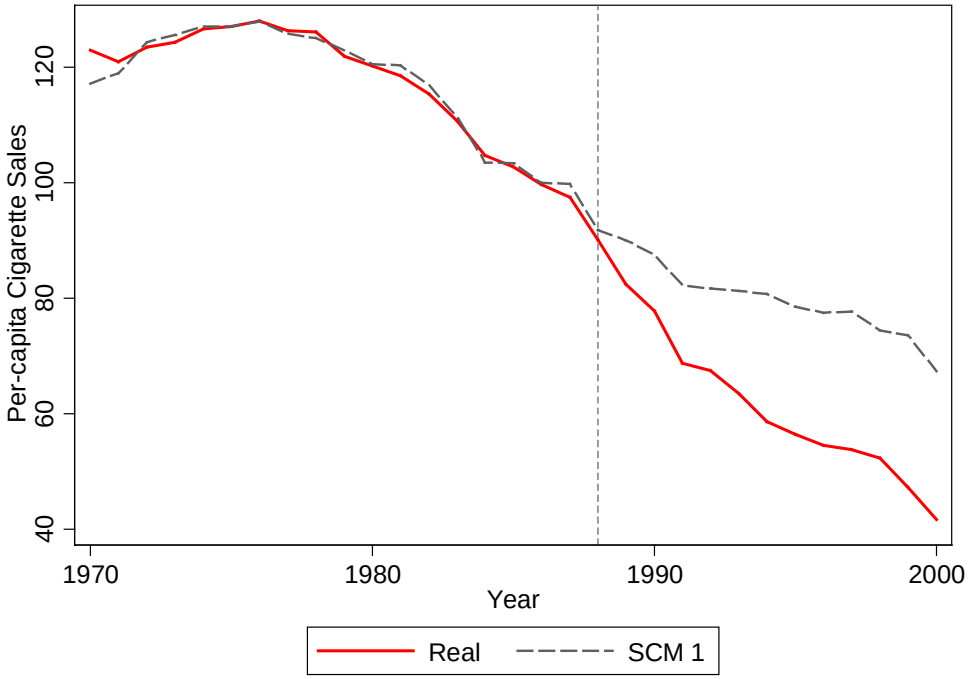


Figure 3- 5 Real-Synthetic Trajectories of California Intervention 1989 SCM 2

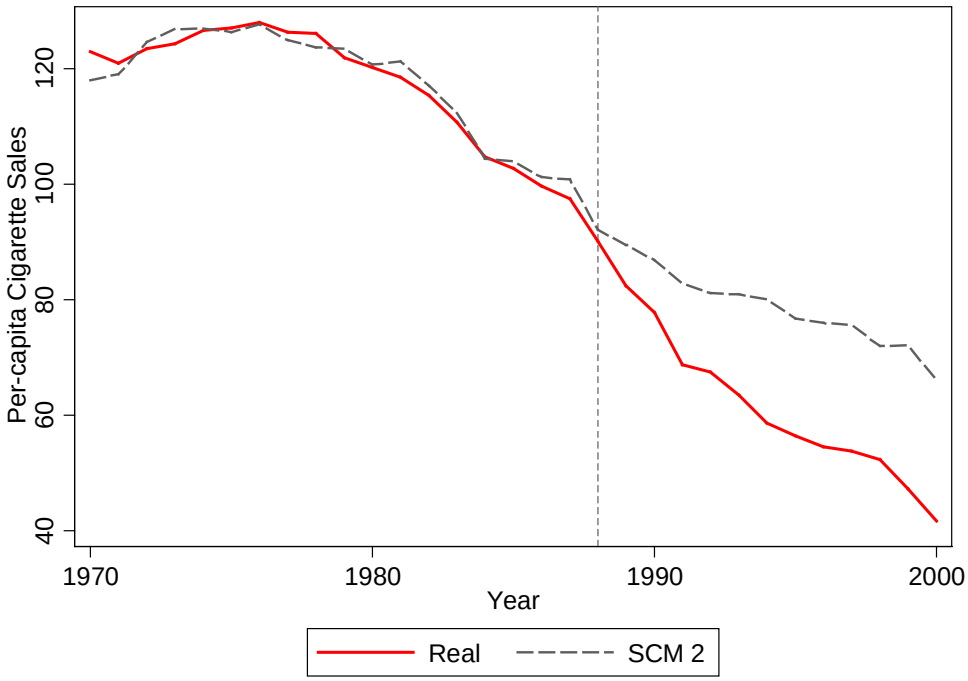
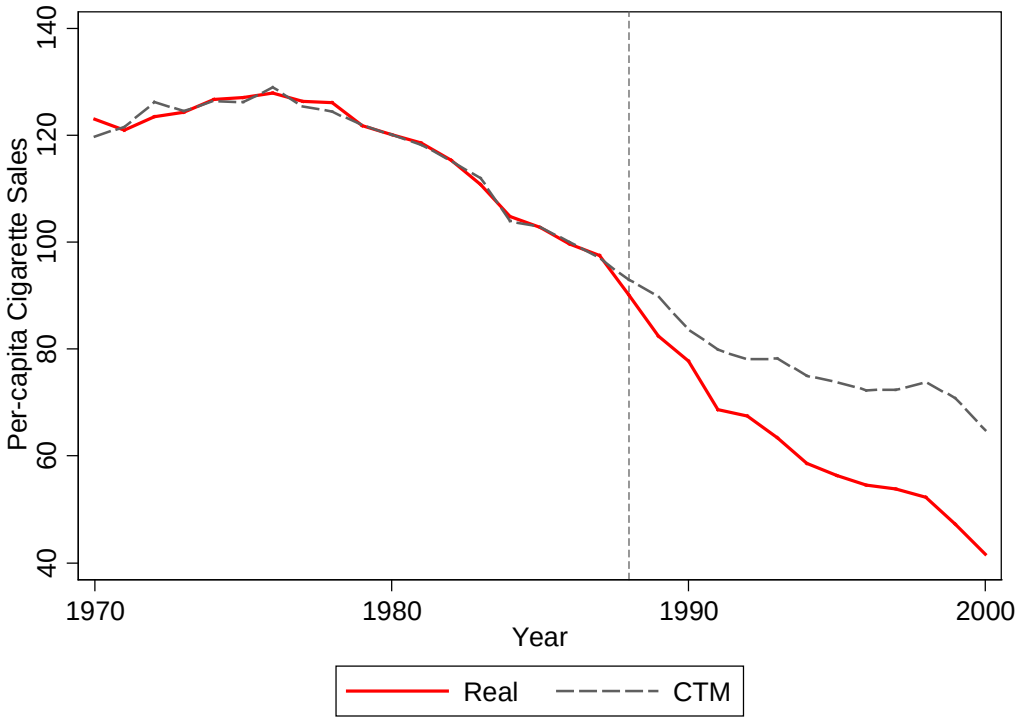


Figure 3- 6 Real-Synthetic Trajectories of California Intervention 1989 CTM



5. CONCLUSIONS

One major focus of this chapter is to conduct a robustness test for major conclusions obtained from prior chapters to the effect of pasteurization ordinances. Using SCM, I found a significant treatment effect in Chicago's 1916 pasteurization ordinance. To verify SCM results, I use a linear regression based algorithm which was introduced by HCW (2012). Instead of using pre-intervention outcomes and non-outcome covariates, HCW conducted their estimation using outcome variables in form of cross-sectional time series (CTM). Unlike regular OLS regression, HCW applied a subset selection process to avoid the danger of over-fitting. In other words, there are only some units used to construct the synthetic counterfactual, although better fitting can be achieved by using more units as regressors.

This paper first uses CTM to re-estimate effects of Chicago's 1916 pasteurization ordinance. According to HCW, we can use outcome predictors only if information of other covariates has been "embedded" into outcomes. In this case study, CTM results are very close to results SCM models, and CTM trends are akin to SCM trends. In addition, the estimated real-synthetic gaps are significant after intervention, in both SCM and CTM models. Thus, CTM results confirm prior findings regarding the treatment effect of Chicago's 1916 ordinance.

Then, I use another dataset, the 1989 California cigarette sales tax, to make an "out-of-sample" robustness test for CTM. Similar to the procedure above, CTM results are comparable with SCM specifications. Synthetic trends generated from CTM, SCM 1 (nested) and SCM 2 (regular) are similar. The estimated treatment effects are

similarly significant in all three models. CTM could be used as an alternative for SCM models.

Overall, CTM and SCM share some similarities. First, the treatment effect is estimated with one treated unit and a set of control units. Second, both CTM and SCM need an algorithm to choose a subset of control units to construct the counterfactual. However, in CTM, this procedure is completed as predictor selection. In addition, the two studies in this paper also highlight one concern in SCM applications, the "volatility condition". In principle, SCM prefers non-volatile outcomes to construct the synthetic version. For example, the dependent variable of California's cigarette sale is not as volatile as Chicago's mortality rates. As a result, California's pre-intervention real-synthetic fitting is better than Chicago.

In sum, this paper confirms robustness of prior SCM results using CTM. In addition, it provides evidence on the performance of CTM estimators. Some technical issues are discussed in this paper as well, which might aid practitioners in handling predictor selection, volatility of outcomes and standard errors calculations.

APPENDIX

Proposition 1: Embedded information

The treatment effect estimation of HCW (2012) relies on a fundamental assumption that the information provided by common factors is embedded in observed outcomes. We provide a case to show how this proposition can be applied in practice. Considering the time-series properties of outcomes, its data generating process (DGP) is specified as,

We assume $y_t$is a process depends on its lagged terms and the common factors of X_t. Also, X_t satisfies an autoregressive process.

$$y_t = \rho y_{t-1} + \beta X_t + u_t \quad \text{for } t = 1, \cdots, T+P \tag{A1}$$

$$X_t = \gamma X_{t-1} + v_t \quad \text{for } t = 1, \cdots, T+P \tag{A2}$$

Assumption 1: $\{ u_t \}$ are iid random variables with a white noise process, $u_t \sim WN(0, \sigma_u^2)$

Assumption 2: $E(u_t|X_t) = 0$

Assumption 3: $\{v_t\}$ follow a white noise process, $v_t \sim WN(0, \sigma_v^2)$

Assumption 4: $|\rho| < 1$, $|\gamma| < 1$

From (2), apply the iterative process in (1),

$$y_{t-1} = \rho y_{t-2} + \beta X_{t-1} + u_{t-1}$$

Substituting X_t with terms of its lagged term, y_t can be expressed as

$$y_t = \rho y_{t-1} + \beta(\gamma X_{t-1} + v_t) + u_t$$

$$X_{t-1} = \frac{1}{\beta}(y_{t-1} - \rho y_{t-2} + u_{t-1})$$

$$y_t = \rho y_{t-1} + \beta[\gamma \frac{1}{\beta}(y_{t-1} - \rho y_{t-2} - u_{t-1}) + v_t] + u_t$$

$$y_t = \rho y_{t-1} + \gamma y_{t-1} - \gamma\rho y_{t-2} + \beta\gamma v_t - \gamma u_{t-1} + u_t$$

$$y_t = (\rho + \gamma) y_{t-1} - \gamma\rho y_{t-2} + (\beta\gamma v_t - \gamma u_{t-1} + u_t)$$

The representation of y_t is now as a combination of its lagged values and error terms. It can be estimated by (A3).

$$E(y_t \mid y_{t-1}, y_{t-2}) = (\rho + \gamma) y_{t-1} - \gamma\rho y_{t-2} \qquad \text{(A3)}$$

REFERENCES

Abadie, A., A. Diamond, and J. Hainmueller. “Comparative Politics and the Synthetic Control Method.” *American Journal of Political Science* (online) APR 2014.

Abadie, A., A. Diamond, and J. Hainmueller. “Synthetic Control Method for Comparative Case Study: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105, no. 490 (2010): 493–505.

Abadie, A., and J. Gardeazabal. “The Economic Costs of Conflict: A Case Study of the Basque Country.” *American Economic Review* 93, no.1 (2003): 112–32.

Abadie, A., “Using Synthetic Controls to Evaluate an International Strategic Positioning Program in Uruguay” [Online], Available at: http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=37312864

Bai, J., and S. Ng. “Determining the number of factors in approximate factor models.” *Econometrica* 70, no. 1 (2002): 191-221

Beaver, M. W. “Population, infant mortality and milk.” *Population Studies* 27, no. 2 (1973): 243-254.

Billmeier, A., and T. Nannicini. “Assessing economic liberalization episodes: A synthetic control approach.” *Review of Economics and Statistics* 95, no. 3 (2013): 983-1001.

Bohn, S., M. Lofstrom, and S. Raphael. “Did the 2007 Legal Arizona Workers Act reduce the state's unauthorized immigrant population?.” *Review of Economics and Statistics* 96, no. 2 (2014): 258-269.

Card, D. (1990), “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, 44, 245–257

Card, D., and Krueger, A. B. (1994), “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *American Economic Review*, 84, 772–793

Cutler, D., and G. Miller. “The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.” *Demography* 42, no.1 (2005): 1–22.

Dube, A., and B. Zipperer. “Pooled Synthetic Control Estimates for Recurring Treatments: An Application to Minimum Wage Case Studies.” Unpublished paper (2013)

Giannone D., M. Lenza, L. Reichlin 2010. Business cycles in the euro area. In Europe, and the Euro, Alesina A, Giavazzi F (eds). National Bureau of Economic Research: Cambridge, MA; 141-167

Hall, Carl W., and M. G. Trout. *Milk Pasteurization.* Westport, CT: AVI Pub. Co, 1968.

Hastie, Trevor, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The Elements of Statistical Learning*. Vol. 2, no. 1. New York: springer, 2009

Hsiao C., Ching S. H. and Wan K . 2012. A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics* 27: 705-740

Hsiao, C., and S. K. Wan. "Is there an optimal forecast combination?." *Journal of Econometrics* 178 (2014): 294-309.

James, G., D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. New York: Springer, 2013

Kiesel K. and Villas-Boas S. B. 2013. Can information costs affect consumer choice? Nutritional labels in a supermarket experiment, *International Journal of Industrial Organization* 31: 153-163

Lee, K-S. "Infant mortality decline in the late 19th and early 20th centuries: the role of market milk." *Perspectives in biology and medicine* 50, no. 4 (2007): 585-602.

Meckel R. A., *Save the Babies: American public health reform and the prevention of infant mortality*, 1850-1929, Johns Hopkins University Press, 1990

Rodrik, D. *One economics, many recipes: globalization, institutions, and economic growth*. Princeton University Press, 2008.

Selitzer R . 1976. *The dairy industry in America*. Magazines for Industry: New York

Vögele, Jörg, and Wolfgang Woelk. "Public Health and the Development of Infant Mortality in Germany, 1875–1930." *The History of the Family* 7, no.4 (2002): 585–99.

Wolf, J. H. "Low Breastfeeding Rates and Public Health in the United States." *American Journal of Public Health* 93, no. 12 (2003): 2000–10.

CHAPTER IV.

HEALTH IMPACT OF VOLUNTARY ADOPTION OF PASTUERIZATION IN THE EARLY 20TH CENTURY: A FIXED-EFFECT ESTIMATION

Huiqiang Wang

Department of Environmental and Natural Resource Economics,

University of Rhode Island, Kingston, RI, USA 02881

July 2015

Abstract

More than mandatory pasteurization ordinances, voluntary pasteurization noticeably increased in the early 1920s across cities in the United States. Using a two-period panel 1921-1924, this study measures the health impacts of variations of extent of pasteurization. Empirically, I choose the Fixed-Effects model to control unobserved intra-city variations. With respect to influential observations, I use robust estimators to validate least squares estimations. Compared with OLS estimate, robust estimates of the coefficients are smaller in absolute value. But their standard errors are even lower. In sum, my FE regressions also support the positive health effect of pasteurization.

Keywords: Public Health, Pasteurization, Robust Panel Estimators

1. INTRODUCTION

Prior chapters discussed causal health effects of mandatory pasteurization ordinances in the United States in the 1910s. Using annual data and comparing sample averages, Difference-in-Difference (DD) models suggested significant effects associated with city ordinances in 1916. However, results from synthetic control methods (SCM) had different results, as it considers unit-specific and time-variant factors. Nonparametric inferences to results of multiple treated units SCM (MTSCM) suggested the treatment effects were not statistical significant at the group level. Although historical evidences told us how milk pasteurization was critical to the drop of children diarrhea mortality, MTSCM results indicated significant health benefits only exists in some cities.

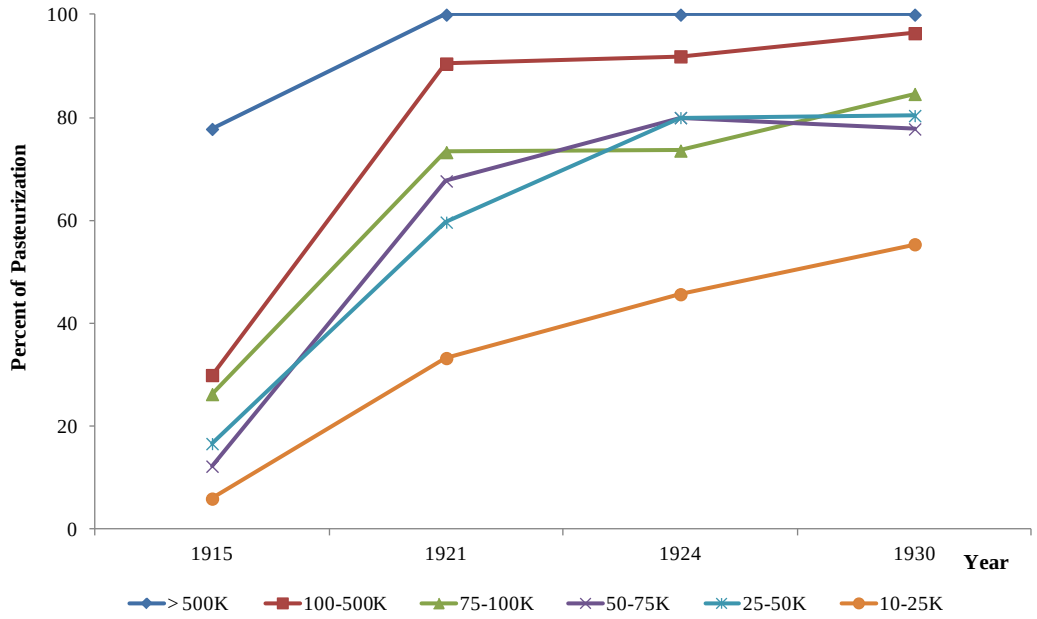
Motivated by prior puzzling results, this chapter aims to clarify the role of milk pasteurization using an alternative approach. Instead of focusing on mandatory city ordinances, this paper measures health impacts of voluntary pasteurization in the early 20^{th} century. Historical facts suggested commercial pasteurization experienced two waves spread in the United States. The first wave occurred in large cities in the early 1910s. Many large cities either recommended or requested pasteurization to most milk sold in the market (Straus 1917). The second wave happened in the late 1910s to the early 1920s when pasteurized milk was increasingly available in small cities. This wave was even stronger, especially in cities without ordinances (Ayers 1922, 1926).

For example, Figure 1 plots changes of the extent of pasteurization in cities by their population. In large cities (population > 500,000), pasteurization leveled off in the early 1920s at almost 100 percent. Before that, many large cities issued mandatory ordinances and required all milk (except certified milk) to be pasteurized before sale.

In this graph, we see that sharp growth of pasteurization occurred in the group with population of 100,000 to 500,000, the group of 75,000 to 100,000, the group of 50,000 to 75,000, and the group of 25,000 to 50,000. In small towns (with population lower than 25,000), remarkable increases continued from the middle 1910s to the early 1930s. Overall, pasteurization was lower in smaller cities. In cities with population over 100,000, the extent was over 90%. But the extent was lower than 60% in cities with population lower than 25,000. Focusing on the late 1920s, Figure 2 shows a similar trend, namely noticeable increases came from small cities. Figure 3 compares extents of pasteurization across regions. On average, the South and the Central South had lower pasteurization. New England, Middle Atlantic and East North Central had the highest shares.

The above discussions highlighted variation of pasteurization across cities. So a new perspective for the health impact of pasteurization is to examine health outcomes of the spread of this technology. To reveal the whole health picture, this paper estimates the relationship of health outcomes to the share of milk pasteurized in the early 1920s.

Figure 4- 1 Increase of Pasteurization Across Cities (by population)



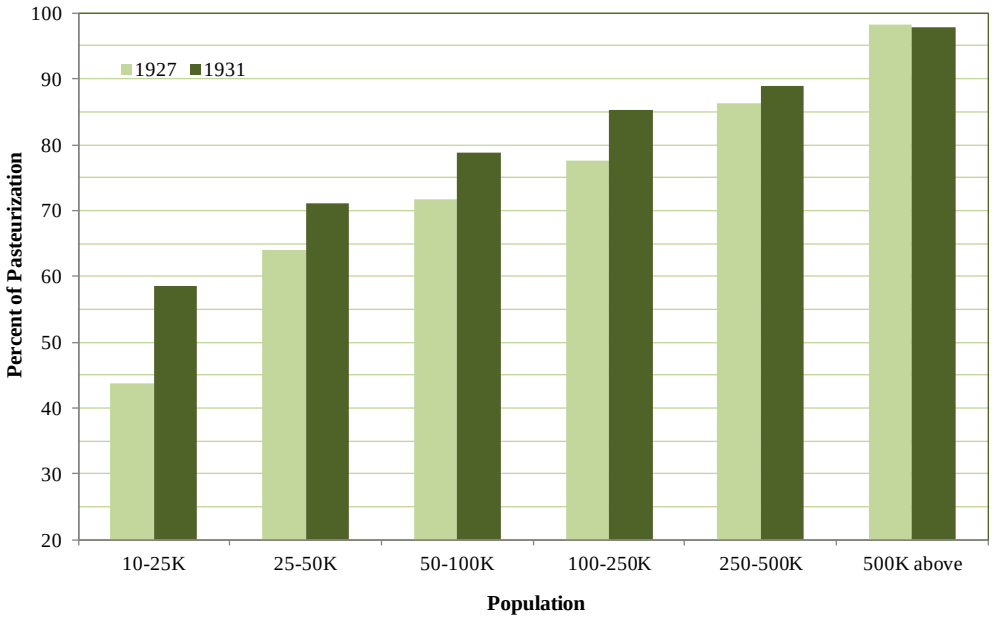
Source: *Present status of the pasteurization of milk*, Bulletin No. 342. USDA (Ayers 1922, 1926, 1932)
Note: 1,000 people as 1 K

Table 4- 1 Summary Extent of Pasteurization in Cities 1921, 1924 and 1930

City Population	Number of cities reporting			Number of cities without milk pasteurized			Average % of milk pasteurized		
	1921	1924	1930	1921	1924	1930	1921	1924	1930
> 500K	12	9	11	0	0	0	95.0	98.1	97.1
100-500K	42	37	56	0	0	0	72.0	81.7	84.9
75-100K	15	19	13	0	0	0	68.0	66.6	81.5
50-75K	29	25	37	5	0	0	65.0	66.6	72.2
25-50K	55	60	56	7	2	0	58.0	67.0	73.1
10-25K	77	105	92	49	21	6	51.0	42.5	52.1
<10K	36	73	79	52	20	44	53.0	33.0	27.1
Total	266	328	344	113	43	50	----	----	----

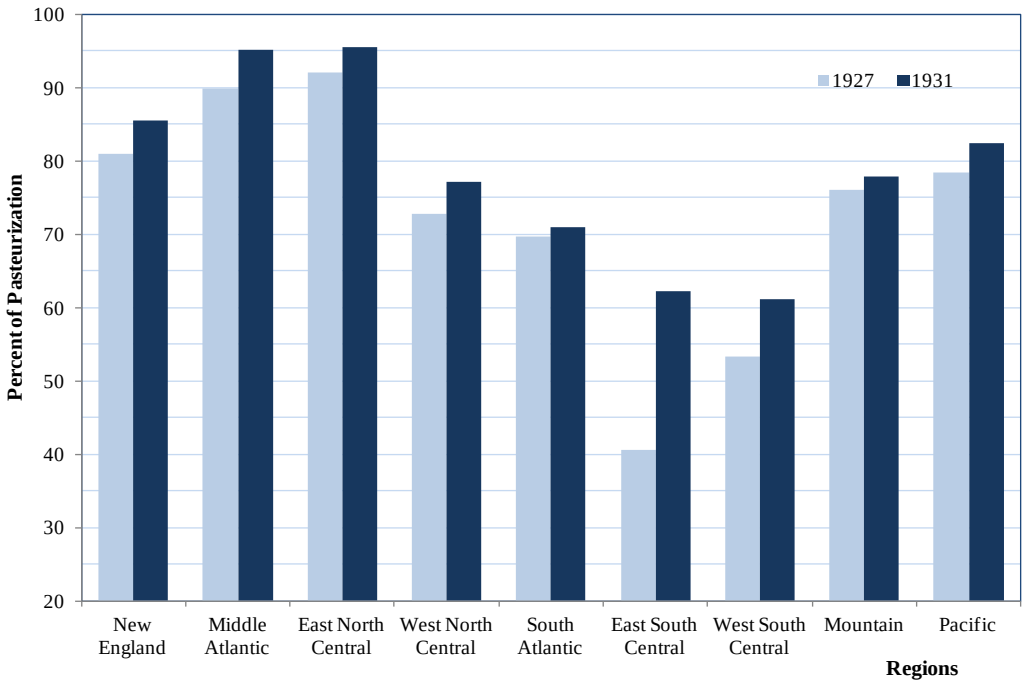
Source: *Present status of the pasteurization of milk*, Bulletin No. 342. USDA (Ayers 1922, 1926, 1932)
Note: 1,000 people as 1 K

Figure 4- 2 Percentage of Pasteurization in Cities (by Population)



Source: Frank and Moss, *The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931*. US Public Health Service Note: 1,000 people as 1 K

Figure 4- 3 Percentage of Pasteurization in Cities (by Region)



Source: Frank and Moss, *The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931*. US Public Health Service Note: 1,000 people as 1 K

2. DATA AND SPECIFICATIONS

2.1 Data

As mentioned, this paper centers on the share of pasteurization across cities in 1921-1924 in the United States. Rationale to use this period is three-fold. First and foremost, data availability, some agencies began to collect the share of pasteurization from the 1920s. For example, the Department of Agriculture (USDA) did surveys in 1921, 1924 and 1930. Another agency, the US Public Health Service (USPHS) made a similar series of survey but more from public health concerns in 1927, 1931 and 1936. Data were available from USDA in years of 1921, 1924 and 1930. USPHS published a survey in 1931. Second, as discussed above, pasteurization noticeably increased in the early 1920s. Thus, this period is preferable to observe how the extent of pasteurization was correlated with health outcome changes. In many cities, the earlier or the later variations were smaller, as Figure 1 and Figure 2 indicate. Third, for the empirical strategy, the Fixed-Effect (FE) model in this study, it is desirable to have a shorter period for estimation, especially because some covariates were not available, so they are aggregated into the fixed effect term. Thus, it is not rigorous to assume these factors kept unchanged in a longer period. The period 1921-1924 is the shortest span within my data availability.

The USDA data are obtained from Ayers (1922, 1926). In the 1921 survey, 266 cities reported the share of pasteurization. And 285 cities did in 1924. Table 1 summarized extent of pasteurization. Interestingly, pasteurization was not always increasing. In some cities, the share of milk pasteurized also dropped. To this analysis, it is good to use variations at different directions. For the health outcome, I

use the same children diarrhea mortality rate as before. Combining pasteurization shares and mortality rates, I got a balanced panel with 109 cities in two periods.

2.2 Specification

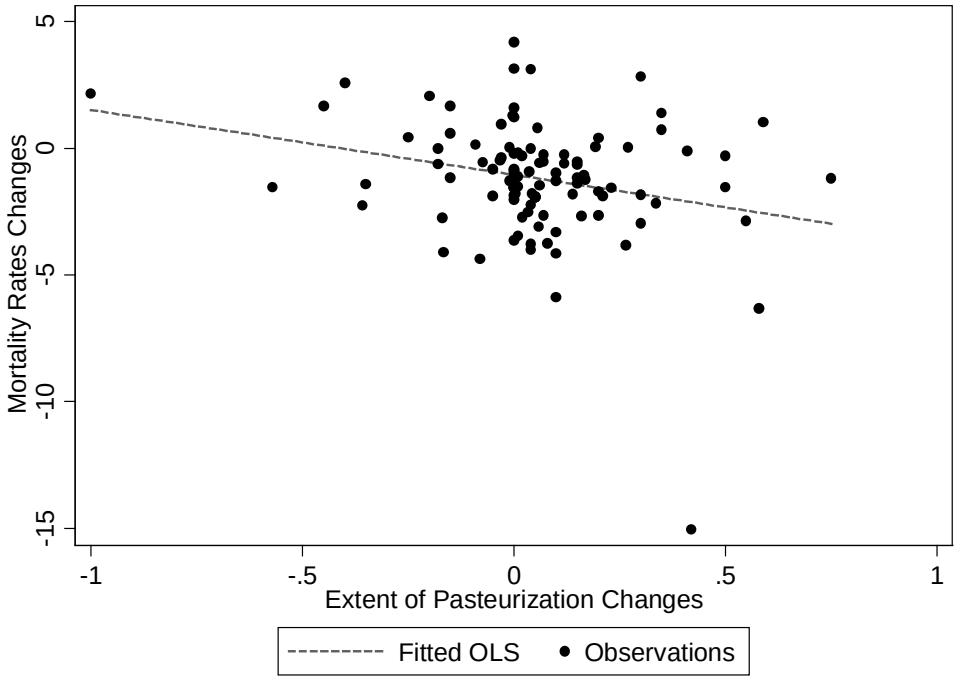
To confirm the health impacts, I need to find a more general trend across units which can support SCM and MTSCM findings. Thus, I estimate the relationship of health outcomes to the share of milk pasteurized. As a continuous variable, the marginal effect of variations in the extent of pasteurization is more meaningful to the health effect of pasteurization in a larger sample. These estimates will add new evidence regarding voluntary pasteurization measures to the public health discussion.

Nonetheless, there are some econometric concerns. First, omitted variable bias exists. To address this concern, I use FE models to control unobserved factors, and then estimate health outcomes associated with variations of pasteurization. Surveys about the extent of pasteurization were available in 1921 and 1924. The short panel about these dates is fortunate in view of the rapid pace of social-economic and public health in inter-war years. Unobserved intra-city variation over three years was probably minor compared to the cross-city variation captured by the FE model.

Second, my FE model faces the challenge of influential points or outliers as well. An observation is influential if “its omission from the sample induces a substantial change in a parameter of interest” (Hansen 2014). According to Hampel (1973), influential deviations are generated from data rounding and grouping, random gross errors, and approximations of assumed models (with central limit theorem). Simply speaking, influential observations are from the other data generating processes

(DGPs). This problem is not rare for historical data. Regular Ordinary Least Squares (OLS) and FE models are very sensitive to outliers from other DGPs. They have large effect on the mean and drag it towards them (Rousseeuw and Leroy 1987). Graphically, it is clear that an influential observation can tilt the LS fitted line toward it, as Figure 4. For remedy, I will use robust estimator for my FE model.

Figure 4- 4 Outliers Detection in Fixed-Effects Model: Changes in the Extent of Milk Pasteurized and Mortality Rates



My FE model is specified as (1). There are 109 cities are included.[65] i = 1, … , 109 and t = 0 for 1921 and 1 for 1924.

$$Y_{it} = \alpha_i + \delta_t + \mathbf{X}'_{it}\theta + \varepsilon_{it} \qquad (1)$$

$$(Y_{it} - \overline{Y}_{it}) = (\alpha_i - \overline{\alpha}_i) + (\mathbf{X}_{it} - \overline{\mathbf{X}}_{it})\theta + (\varepsilon_{it} - \overline{\varepsilon}_{it})$$

[65] This sample includes both cities with and without mandatory pasteurization ordinances.

In the model, Y_{it} is the mortality rate of city i in year t. City fixed-effect is α_i, and year fixed-effect is δ_t. Covariates vector $\mathbf{X}$ includes share of pasteurization and population (log values). In this study, I did not include decennial demographic and income variables as covariates. Population is used to proxy the difference between large and small cities. Changes in other variables are included in the city fixed effect term. Since only two periods are used in the FE model, it can also be transformed as the First Differenced (FD) form in (2).

$$\Delta Y_{it} = \Delta \mathbf{X}'_{it}\theta + \Delta \varepsilon_{it} \quad (2)$$

According to Rousseeuw and Leroy (1987), there are three categories of influential points, including good leverage points, bad leverage points and vertical outliers. First, vertical outliers are outlying in vertical values, but still within the space of explanatory variables. They affect the intercept of LS estimators. Second, good leverage points are close to the regression line but outlying the space of explanatory variables. They have no direct effect to the LS coefficients but lead to inflated standard errors. Third, bad leverage points are outliers in the spaces of both dependent and explanatory variables. They affect LS estimations in both intercept and slope (Verardi and Croux 2009). Figure 4 plots within-group variations of the extent of pasteurization and mortality rate changes, which is the major focus of my FE model. [66]

Motivated by inefficient LS estimation in a contaminated sample, robust estimators have been developed to control the influence of outliers. Robust estimators are insensitive to small deviations from the assumptions made (Huber 1996). These estimators give results with small sampling variances, and are robust to small

[66] Graphically, it follows the pattern of Verardi and Croux (2009). Definition of the types of outliers is in Rousseeuw and Leroy (1987).

deviations from the assumed models; the effects of larger deviations from the assumed models are within a reasonable range (Huber and Ronchetti 2009). Besides regular OLS estimation, I also include four robust estimators (L-estimator, M-estimator, S-estimator, MM-estimator) for FE (or FD) model in (1). Details of robust estimators are available in the Appendix.

Table 4- 2 Extents of Pasteurization and Mortality Rates, 1921-24

Specifications	% milk pasteurized	Population (log)	Year dummy	Max mortality [1] rate drop %
LS-estimator	-2.409**	3.306	-1.277***	-60.2%
	(1.215)	(2.432)	(0.238)	
L-estimator	-1.992***	0.461***	-0.429***	-49.8%
	(0.532)	(0.114)	(0.084)	
M-estimator (95%)[2]	-2.042***	0.443***	-0.362***	-51.1%
	(0.518)	(0.096)	(0.077)	
M-estimator (70%)[3]	-1.975***	0.448***	-0.435***	-49.4%
	(0.404)	(0.069)	(0.064)	
S-estimator	-1.573***	0.538***	-0.561***	-39.3%
	(0.278)	(0.183)	(0.067)	
MM-estimator	-1.550***	0.511***	-0.432***	-38.8%
	(0.479)	(0.118)	(0.075)	
Observations	218	(all models)		

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
1. The % of mortality rate decline compared with values in 1921, if pasteurization increased 100%.
2. M-estimator with 95% Gaussian Efficiency
3. M-estimator with 70% Gaussian Efficiency
4. S-estimator with fixed effects
5. MM-estimator with fixed effects
6. Standard errors of LMS-estimator and LTS-estimator are not reported (N.A.).

3. RESULTS

The results in Table 2 indicate the marginal effect of pasteurization on the conditional mean of children morality rate, holding other variables constant. The LS-estimator suggests that one percentage increases in the share of milk pasteurized decreases the mortality rate by 0.02409 (-2.409/100). Coefficients of L-estimator, M-estimator (95% Gaussian efficiency), M-estimator (70% Gaussian efficiency), S-estimator, and MM-estimator are -0.01992, -0.02045, -0.01975, -0.01573 and -0.01550. Compared with OLS estimate, robust estimates of the coefficients of interest are smaller in absolute value. But their standard errors are even lower. Thus, the coefficients are statistically significant at more rigorous levels.

In sum, my FE regressions also support the positive health effect of pasteurization. They provide more cross-unit evidence for individual specific case studies in SCM and MTSCM analysis. From a public health perspective, we can conclude that in the early 1920s, increases in the share milk pasteurized were associated with decreases in children diarrhea mortality rates. From a public policy perspective, we can infer that mandatory ordinances could increase pasteurization, particularly in smaller cities.

4. CONCLUSIONS

Using a panel data set of the extent of pasteurization in 1921 to 1924, this study estimated the relationship between health outcomes and extent of pasteurization across cities using a fixed-effects regression model. With respect to the concern of influential points in the data, I choose both the OLS estimator and robust estimators to measure the health impact of changes in the share of milk pasteurized. My results indicate the increases of pasteurization were significantly correlated with the drop of child diarrhea mortality rates. Unlike the OLS estimate, robust estimates of the coefficients of interest are smaller in absolute value, but their standard errors are even lower. Thus, the coefficients are statistically significant at more rigorous levels. In sum, my FE regressions also support the positive health effect of pasteurization. This study could be used to cross-validate evidence obtained from prior chapters.

APPENDIX

A1. Consequences of An Influential Point to OLS Estimation

This section is largely with reference to an online lecture of Hansen (2014).[67] It aims to illustrate how an outlier or influential point would affect regular OLS estimations.

In regular OLS framework, we have the coefficient of one explanatory to estimate.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and } \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\begin{aligned} &\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \ \text{ with } \sum_{i=1}^{n} x_i x_i = \mathbf{X}'\mathbf{X} \text{ and } \sum_{i=1}^{n} x_i y_i = \mathbf{X}'\mathbf{y} \\ &\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{aligned} \tag{A1}$$

Suppose there is one outlier *i* in sample, then leave-one-out (*i*) OLS estimator is $\hat{\beta}_{(-i)}$

$$\begin{aligned} \hat{\beta}_{(-i)} &= \left(\frac{1}{n-1} \sum_{j \neq i} x_j x_j' \right)^{-1} \left(\frac{1}{n-1} \sum_{j \neq i} x_j y_j \right) \\ &= \left(\mathbf{X}_{(-i)}' \mathbf{X}_{(-i)} \right)^{-1} \left(\mathbf{X}_{(-i)}' y_{(-i)} \right) \end{aligned} \tag{A2}$$

Here, we define $\tilde{y}_i = x_i' \hat{\beta}_{(-i)}$ and $\tilde{e}_i = y_i - \tilde{y}_i$

$$\begin{aligned} &\hat{\beta} - \hat{\beta}_{(-i)} = (1 - h_{ii})^{-1} (\mathbf{X}'\mathbf{X})^{-1} x_i \hat{e}_i = (\mathbf{X}'\mathbf{X})^{-1} x_i \tilde{e}_i \\ &\text{where the leverage is } h_{ii} = x_i' (\mathbf{X}'\mathbf{X})^{-1} x_i \end{aligned} \tag{A3}$$

The difference between our full sample estimated value ($\hat{y}_i$) and leave-one-out prediction ($\tilde{y}_i$) is defined in the next equation.

$$\hat{y}_i - \tilde{y}_i = x_i' \hat{\beta} - x_i' \hat{\beta}_{(-i)} = x_i' (\mathbf{X}'\mathbf{X})^{-1} x_i \tilde{e}_i = h_{ii} \tilde{e}_i \tag{A4}$$

An influential point or outlier has large $|h_{ii}\tilde{e}_i|$. A leverage observation is defined as an point with large values of h_{ii}. But a leverage point is not necessarily an outlier. The latter also requires large values in $\tilde{e}_i$ (Hansen 2014).

[67] Hansen (2014)

A2. Robust Estimators

To control the influence of outliers, robust regression estimators have been developed over past decades. In this paper, robustness is defined as "small deviations will not significantly affect the conclusions drawn from the data" (Stuart 2011).[68] Robust estimators are insensitive to small deviations from the assumptions are made (Huber 1996). Results generated from this class of estimator have small sampling variances, and are robust to small deviations from the assumed models. Or the effects of larger deviations from the assumed models are within a reasonable range (Hubert and Ronchetti 2009).

Performances of robust estimators are empirically evaluated by fraction of breakdown points and their relative efficiencies. Breakdown point (BDP) measures the resistance to outliers. BDP is the smallest share of the "contaminated data" which can cause the estimator to break down or cannot represent the real trend in the uncontaminated data (Stuart 2011). Formally, it is formally defined as following. T is the regression estimator, Z is a sample of n data points, and $Z'$is the sample with m outliers and n points in total. $T(Z) = \hat{\beta}\ and\ T(Z') = \tilde{\beta}$

The maximum contamination effect is

$$\text{Effect}(m; T, Z) = \sup_{Z'} \left\| T(Z') - T(Z) \right\| \tag{A5}$$

BDP can mathematically be defined as

[68] Stuart (2011) Robust Regression

$$BDP(T,Z) = \min\left\{\frac{m}{n}: \text{Effect(m;T,Z) is finite}\right\} \quad \text{(A6)}$$

In Appendix, I illustrate why OLS estimator is sensitive to outliers. Its BDP is 1/n, as just one leverage point can break down regular OLS estimation. When the sample size increases, the BDP of OLS estimators will be 0%. In contrast, robust estimators have much higher BDP (Stuard 2011). A rule of thumb is good robust estimators have BDP as high as 50%. If contaminated data are over 50%, researchers cannot identify this sample is good or not good (Binaco et al. 2005).[69]

A second criterion for robust regression is relative efficiency (*RE*), as defined below. Suppose we have two estimators $\hat{\beta}_1$and $\hat{\beta}_2$, $\hat{\beta}_1$ is the efficient one and $\hat{\beta}_2$is the less efficient one. And β is a population parameter. Then, *RE* is the variance ratio of these two estimators (Andersen 2008). [70]

$$\text{Relative Efficiency } (\hat{\beta}_1, \hat{\beta}_2) = \frac{E[(\hat{\beta}_1 - \beta)(\hat{\beta}_1 - \beta)']}{E[(\hat{\beta}_2 - \beta)(\hat{\beta}_2 - \beta)']} \quad \text{(A7)}$$

In practice, RE is used as Gaussian Efficiency which is calculated with normal errors. When errors of $\hat{\beta}_1$are normally distributed A7, then $0 < RE \leq 1$. For researchers, high BDP and high RE are desirable. However, in most cases, we cannot get an estimator like that. Instead, we need an estimator which can balance BDP and RE. With a series of iterative algorithms, robust estimators achieve their robustness by modifying the loss function.

[69] Bianco, Ben, and Yohai (2005): 511-528
[70] Andersen (2008)

Regular LS estimators are obtained by minimizing squared residuals, which tends to give excessive importance to large residuals[71].

$$\widehat{\theta}_{LS} = \arg\min_{\theta} \sum_{i=1}^{n} r_i^2(\theta);\ r_i = y_i - \hat{\alpha} - x_i'\widehat{\theta} \quad \text{for } 1 \le i \le n \tag{A8}$$

In contrast, the median or L-estimator minimizes the sum of absolute value of residual r_i. L-estimator is robust to vertical outliers but has a low Gaussian efficiency.

$$\widehat{\theta}_{L} = \arg\min_{\theta} \sum_{i=1}^{n} \left| r_i(\theta) \right| \tag{A9}$$

Based on Maximum Likelihood algorithm, M-estimator is resistant to vertical outliers and also has higher efficiency. But it is not robust to bad leverage points. Residuals r_i are standardized by a scale of dispersion (σ). Then it minimizes the loss function $\rho(\cdot)$.

$$\hat{\theta}_{M} = \arg\min_{\theta} \sum_{i=1}^{n} \rho\left\{ \frac{r_i(\theta)}{\sigma} \right\} \tag{A10}$$

The loss function is even, non-decreasing for positive values and less increasing than the square function (Verardi and Croux 2009). As a weighted LS-estimator, the weight w_i of M-estimator is defined in Eq.2.

$$\widehat{\theta}_{M} = \arg\min_{\theta} \sum_{i=1}^{n} \omega_i r_i^2(\theta);\ \text{where } \omega_i = \rho\left\{ \frac{r_i(\theta)}{\sigma} \right\} \cdot \frac{1}{r_i^2} \tag{A11}$$

Unlike M-estimator, S-estimator awards lower weights to large residuals by using a new loss function. It applies a robustly scaled residual ($\hat{\sigma}^S$) to minimize the loss function. S-estimator is robust up to 50% outliers but has a relative low efficiency.

The loss function and S-estimator are in Equations.

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left\{ \frac{r_i(\theta)}{\hat{\sigma}^S} \right\} = b;\ \text{where } b = E[\rho(Z)] \text{ with } Z \sim N(0,1) \tag{A12}$$

[71] Notations follow Verardi and Croux (2009).

$$\hat{\theta}_S = \arg\min_{\theta} \hat{\sigma}^S \{r_1(\theta), r_2(\theta), \ldots, r_n(\theta)\} \tag{A13}$$

MM-estimator combines both S-estimators' high breakdown point and M-estimator's Gaussian efficiency. It is similar to M-estimator but it uses a fixed scale $\hat{\sigma}^S$ to standardize the residuals. First, it uses S-estimator to obtain the scale parameter $\hat{\sigma}^S$ at a break down point of 50%. Next, it assumes the M-estimator and achieves some high Gaussian efficiency by choosing an appropriate ρ function. MM-estimator is defined by Yohai (1987), as A14.

$$\hat{\theta}_{MM} = \arg\min_{\theta} \sum_{i=1}^{n} \rho\left\{\frac{r_i(\theta)}{\hat{\sigma}^S}\right\} \tag{A14}$$

Other robust estimators include LMS-estimator (least median of squares) and LTS-estimator (least trimmed squares), which can be found in Rousseeuw and Leroy (1987).

REFERENCES

Andersen, R. Modern methods for robust regression, 2008, Sage

Ayers, S. H . 1922. The present status of the pasteurization of milk, Bulletin No. 342. United States Department of Agriculture

Ayers, S. H . 1926. The present status of the pasteurization of milk, Bulletin No. 342. revised edition. United States Department of Agriculture

Bianco, A. M., M. G. Ben, and V. J. Yohai. Robust estimation for linear regression with asymmetric errors. *Canadian Journal of Statistics* 33.4 (2005): 511-528

Frank, L. C. and Moss, F. J . 1931. The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931. US Public Health Service

Hansen, B.E. 2014. Econometrics, online www.ssc.wisc.edu/~bhansen, available July, 2015

Huber, P. J. Robust statistical procedures: SIAM. 1996

Huber, P. J. and E. M. Ronchetti, Robust Statistics. 2nd edition, 2009 Hoboken: Wiley

Rousseeuw P. J., and A. M. Leroy, Robust Regression and Outlier Detection, New York: Wiley, 1987

Stuart, C., 2011. Robust Regression, online http://www.maths.dur.ac.uk/Ug/projects/highlights/CM3/Stuart_Robust_Regression_report.pdf, available July, 2015

Veradi V. and C. Croux, Robust regression in Stata, *The Stata Journal*, 2009, 3 (9): 439-453

Vogel A. and J. Wagner, Robust estimates of exporter productivity premia in German business services enterprises, *Economic and Business Review*, 2011, (13) 1-2: 7-26

BIBLIOGRAPHY

Abadie, A., "Using Synthetic Controls to Evaluate an International Strategic Positioning Program in Uruguay" [Online], available July 2015 http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=37312864

Abadie, A., A. Diamond, and J. Hainmueller. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* (Article first published online: 23 APR 2014).

Abadie, A., A. Diamond, and J. Hainmueller. "Synthetic Control Method for Comparative Case Study: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105, no. 490 (2010): 493–505.

Abadie, A., and J. Gardeazabal. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93, no.1 (2003): 112–32.

Alvord H. E., and R. A. Pearson. "The Milk Supply of Two Hundred Cities and Towns." U.S. Department of Agriculture, Bureau of Animal Industry, Bulletin, 46, Washington, DC: 1903.

Andersen, R. Modern methods for robust regression, 2008, Sage

Atkins Peter J. The Pasteurization of England: The Science, Culture and Health Implications of Milk Processing, 1900–1950. In *Food, Science, Policy and Regulation in the Twentieth Century: International and Comparative Perspectives*, edited by David F. Smith and Jim Phillips, London: Routledge, 2000: 37-51

Atkins, Peter J. *Mother's milk and infant death in Britain, circa 1900–1940. Anthropology of Food* [Online]. Available at: http://aof.revues.org/index310.html. 2003.

Ayers, S. H . 1922. The present status of the pasteurization of milk, Bulletin No. 342. United States Department of Agriculture

Ayers, S. H . 1926. The present status of the pasteurization of milk, Bulletin No. 342. revised edition. United States Department of Agriculture

Bai, J., and S. Ng. "Determining the number of factors in approximate factor models." *Econometrica* 70, no. 1 (2002): 191-221

Beaver, M. W. "Population, infant mortality and milk." *Population Studies* 27, no. 2 (1973): 243-254.

Bertrand, M., E. Duflo, and S. Mullainathan. “How Much Should We Trust Differences-in-Differences Estimates?.” *The Quarterly Journal of Economics* 119, no. 1 (2004): 249-275.

Bianco, A. M., M. G. Ben, and V. J. Yohai. Robust estimation for linear regression with asymmetric errors. *Canadian Journal of Statistics* 33.4 (2005): 511-528

Billmeier, A., and T. Nannicini. “Assessing economic liberalization episodes: A synthetic control approach.” *Review of Economics and Statistics* 95, no. 3 (2013): 983-1001.

Block, D. “Purity, economy, and social welfare in the progressive era pure milk movement.” Journal for the Study of Food and Society 3, no. 1 (1999): 20-27.

Block, D. R. “Public health, cooperatives, local regulation, and the development of modern milk policy: the Chicago milkshed, 1900–1940.” *Journal of Historical Geography* 35, no. 1 (2009): 128-153.

Bohn, S., M. Lofstrom, and S. Raphael. “Did the 2007 Legal Arizona Workers Act reduce the state's unauthorized immigrant population?.” *Review of Economics and Statistics* 96, no. 2 (2014): 258-269.

Cameron, Colin A., and Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press, 2005.

Card, D. (1990), “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, 44, 245–257

Card, D., and Krueger, A. B. (1994), “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *American Economic Review*, 84, 772–793

Cutler, D., and G. Miller. “The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.” *Demography* 42, no.1 (2005): 1–22.

Czaplicki, A . “Pure Milk Is Better Than Purified Milk Pasteurization and Milk Purity in Chicago, 1908-1916”, *Social Science History* 31: (2007) 411-433

Dahlberg, Arthur C., and Harold S. Adams “Sanitary Milk and Ice Cream Legislation in the United States—A Study of Laws and Ordinances Establishing Sanitary Standards for Milk, Cream, and Ice Cream.” National Research Council, Washington, DC: National Academy of Sciences, 1950.

Davis, V. Effa, “Breast Feeding.” Bulletin: Chicago School of Sanitary Instruction 13 (June 8, 1910): 2.

Dube, A., and B. Zipperer. “Pooled Synthetic Control Estimates for Recurring Treatments: An Application to Minimum Wage Case Studies.” Unpublished paper (2013)

Ericson, John. “Chlorination of Chicago’s Water Supply.” *American Journal of Public Health* 8, no. 10 (1918): 249–51.

Fares, M. and E. Rouviere. “The implementation mechanisms of voluntary food safety systems.” Food Policy 35, no. 5 (2010): 412-418.

Ferrie, J. P., and W. Troesken. “Water and Chicago’s Mortality Transition, 1850–1925.” *Explorations in Economic History* 45, no. 1 (2008): 1–16.

Fogel, Robert W. “Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy1.” *The American Economic Review* 84.3 (1994): 369-395.

Frank, L. C. and Moss, F. J . 1931. The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931. US Public Health Service

Giannone D., M. Lenza, L. Reichlin 2010. Business cycles in the euro area. In Europe, and the Euro, Alesina A, Giavazzi F (eds). National Bureau of Economic Research: Cambridge, MA; 141-167

Gobillon, L., and T. Magnac. “Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls.” (2013) Unpublished Paper.

Haines, Michael R. “American Fertility in Transition: New Estimates of Birth Rates in the United States, 1900–1910.” *Demography* 26, no.1 (1989): 137–48.

Haines, Michael R. “Inequality and Infant and Childhood Mortality in the United States in the Twentieth Century.” *Explorations in Economic History* 48, no. 3 (2011): 418–28.

Haines, Michael R. “Inequality and Childhood Mortality: A Comparison of England and Wales, 1911, and the United States, 1900.” *The Journal of Economic History* 45, no. 04 (1985): 885–912.

Hall, C. W. and Trout G M . 1968. Milk pasteurization. Westport, Conn., Co

Hansen, B.E. 2014. Econometrics, online www.ssc.wisc.edu/~bhansen, available July, 2015

Hastie, Trevor, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The Elements of Statistical Learning*. Vol. 2, no. 1. New York: springer, 2009

Hastings, Justine S. “Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California.” *American Economic Review* 94.1 (2004): 317-328.

Henson, S., and J. Caswell. “Food safety regulation: an overview of contemporary issues.” *Food policy* 24, no. 6 (1999): 589-603.

Hoover, Kevin D. “Causality in economics and econometrics.” The new Palgrave dictionary of economics 2 (2008)

Hsiao C., Ching S. H. and Wan K . 2012. A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics* 27: 705-740

Hsiao, C., and S. K. Wan. “Is there an optimal forecast combination?.” *Journal of Econometrics* 178 (2014): 294-309.

Huber, P. J. and E. M. Ronchetti, Robust Statistics. 2nd edition, 2009 Hoboken: Wiley

Huber, P. J. Robust statistical procedures: SIAM. 1996

Imbens, G. and Wooldridge J., Lecture Notes 10, National Bureau Economic Research (NBER) Summer Institute Econometric Lectures,

James, G., D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. New York: Springer, 2013

Jenkins, Jane E. “Region, Politics, Pasteurization, and the Naturalizing Myth of Pure Milk in 1920s Saint John, New Brunswick.” *Acadiensis* 37, no.2 (2008): 86–105.

Kelly, E, and Clement C E . 1931. Market milk. J. Wiley & Sons, Inc: New York; Chapman & Hall Limited: London

Kiesel K. and Villas-Boas S. B. 2013. Can information costs affect consumer choice? Nutritional labels in a supermarket experiment, *International Journal of Industrial Organization* 31: 153-163

King, G., and L. Zeng. “The Dangers of Extreme Counterfactuals.” *Political Analysis* 14, no. 2 (2006): 131–59.

Lee, Clive H. “Regional Inequalities in Infant Mortality in Britain, 1861–1971: Patterns and Hypotheses.” *Population Studies* 45, no. 1 (1991): 55–65.

Lee, K. S. “Infant Mortality Decline in the Late 19th and Early 20th Centuries: The Role of Market Milk.” *Perspectives in Biology and Medicine* 50 no.2 (2007): 585–602.

Levitt J. W., The Healthiest City: Milwaukee and the Politics of Health Reform, 1996, p. 187

Mann, H. B., and D. R. Whitney. “On a test of whether one of two random variables is stochastically larger than the other.” *The Annals of Mathematical Statistics* (1947): 50-60.

McKeown, T. *The Modern Rise of Population*. New York: Academic Press, 1976.

Meckel R. A., *Save the Babies: American public health reform and the prevention of infant mortality*, 1850-1929, Johns Hopkins University Press, 1990

Miller, Grant. “Women’s Suffrage, Political Responsiveness, and Child Survival in American History.” *The Quarterly Journal of Economics* 123, no. 3 (2008): 1287.

Millward, Robert, and Frances N. Bell. “Economic Factors in the Decline of Mortality in Late Nineteenth Century Britain.” *European Review of Economic History* 2, no. 3 (1998): 263–88.

Olmstead, A. L., and P. W. Rhode. “An impossible undertaking: the eradication of bovine tuberculosis in the United States.” *The Journal of Economic History* 64, no. 03 (2004): 734-772.

Olmstead, Alan L., and Paul W. Rhode. “Not on my Farm! Resistance to Bovine Tuberculosis Eradication in the United States.” *The Journal of Economic History* 67.03 (2007): 768–809.

Phillips, Jim, and Michael French. “State Regulation and the Hazards of Milk, 1900–1939.” *Social History of Medicine* 12, no.3 (1999): 371–88.

Preston, S. H., and M. R. Haines. *Fatal Years: Child Mortality in Late Nineteenth-Century America*. Princeton, N.J.: Princeton University Press, 1991.

Roberston, J. D. *Annual Report of the Department of Health of the City of Chicago for the Years 1911–1918 Inclusive*. Chicago, IL: Chicago. Department of Health, 1919.

Rodrik, D. *One economics, many recipes: globalization, institutions, and economic growth*. Princeton University Press, 2008.

Rousseeuw P. J., and A. M. Leroy, Robust Regression and Outlier Detection, New York: Wiley, 1987

Segerson, K. “Mandatory versus voluntary approaches to food safety.” Agribusiness 15, no. 1 (1999): 53-70.

Selitzer R . 1976. The dairy industry in America. Magazines for Industry: New York

Shapiro, S. S., and M. B. Wilk. “An analysis of variance test for normality (complete samples).” *Biometrika* (1965): 591-611.

Stuart, C., 2011. Robust Regression,
online http://www.maths.dur.ac.uk/Ug/projects/highlights/CM3/Stuart_Robust_Regression_report.pdf, available July, 2015

Veradi V. and C. Croux, Robust regression in Stata, *The Stata Journal*, 2009, 3 (9): 439-453

Vogel A. and J. Wagner, Robust estimates of exporter productivity premia in German business services enterprises, *Economic and Business Review*, 2011, (13) 1-2: 7-26

Vögele, Jörg, and Wolfgang Woelk. “Public Health and the Development of Infant Mortality in Germany, 1875–1930.” *The History of the Family* 7, no.4 (2002): 585–99.

Wagstaff, Adam. “Estimating health insurance impacts under unobserved heterogeneity: the case of Vietnam's health care fund for the poor.” Health economics 19.2 (2010): 189-208. *Winslow, C-E A. Man and Epidemics. Princeton, NJ: Princeton University Press, 1952.*

Winslow C-E A . 1952. Man and epidemics, Princeton University Press. Princeton: New Jersey

Wolf, J. H. “Low Breastfeeding Rates and Public Health in the United States.” *American Journal of Public Health* 93, no. 12 (2003): 2000–10.

Wolfers, Justin. “Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results.” *American Economic Review* 96.5 (2006): 1802-1820.