Infectious Diseases and the Economy

Dissertation
zur Erlangung des akademischen Grades Doctor rerum politicarum
(Dr. rer. pol.)

genehmigte Dissertation
von

Stefan Pichler, M.Sc.

geboren am: 07.12.1985 in Bozen/Italien

Erstgutachter: Prof. Dr. Michael Neugart
Zweitgutachter: Prof. Dr. Carsten Helm

Fachbereich 1 – Rechts- und Wirtschaftswissenschaften
Hochschulkennzahl D17 – Technische Universität Darmstadt 2013

Datum der Einreichung: 06. Februar 2013
Datum der Verteidigung: 20. März 2013

Danksagung

Ich möchte mich bei allen bedanken, die auf das Zustandekommen dieser Arbeit positiven Einfluss genommen haben.

Ein großer Dank geht dabei an meine beiden Doktorväter Prof. Dr. Michael Neugart und Prof. Dr. Carsten Helm für ihre Betreuung. Prof. Dr. Carsten Helm stand mir in den ersten Jahren meiner Dissertation mit seinem Fachwissen, seinen guten Ideen und kritischen Anmerkungen zur Seite. Prof. Dr. Michael Neugart hat mir in der zweiten Hälfte meiner Dissertation mit Rat und Tat zur Seite gestanden. Dabei habe ich sehr viele wertvolle Kommentare erhalten, die zum Gelingen dieser Arbeit beigetragen haben. Beide haben mich stets unterstützt haben für ein angenehmes und produktives Arbeitsumfeld gesorgt.

Des Weiteren möchte ich mich bei allen Mitgliedern der Prüfungskommission Prof. Dr. Volker Caspari, Prof. Dr. Jens Krüger und Prof. Dr. Dirk Schiereck recht herzlich bedanken. Weiterhin möchte ich mich bei allen aktiven und ehemaligen Mitgliedern der Volkswirtschaftlichen Fachgebiete bedanken, die mich während meiner Promotion begleitet haben und mir immer wieder die Gelegenheit gegeben haben über meine Arbeit zu diskutieren. Besonders hervorheben möchte ich dabei Prof. Martin Karlsson, der mich sehr viel unterstützt hat und von dem ich im ökonometrischen und gesundheitsökonomischen Bereich viel lernen konnte. Des Weiteren bin ich meinem ehemaligen Kollegen und Koautor Daniel Römer zu Dank verpflichtet, auf dessen Rückhalt ich stets zählen konnte. Ich möchte mich auch bei Philip Savage dafür bedanken, dass er verschiedene Teile dieser Arbeit durch gründliches Korrekturlesen verbessert hat. Ein weiterer Dank geht an Prof. Dr. Horst Entorf und Therese Nilsson, Ph.D. für ihren Beitrag und für die gute Zusammenarbeit.

Zu guter Letzt möchte ich meiner Familie und meinen Freunden danken. Insbesondere danke ich meinen Eltern, ohne deren Hilfe dieses Projekt nicht realisiert worden wäre. Ein besonderer Dank geht dabei auch an Wenjia Teng, die mir auf persönlicher Ebene eine wichtige Stütze war und mir durch ihr mathematisches Verständnis über mehrere Hürden in dieser Dissertation geholfen hat.

Contents

List of Figures

List of Tables

CHAPTER

1

Introduction

Health and infectious diseases were, are and probably will be an important factor in economic development. The course of economic history has also been affected by plagues. McNeill (1977) notes that the Black Death had enormous economic and social effects. Moreover, Watts (1999) showed that Venice lost its dominance in international trade due to the Black Death because Dutch and English businessmen acquired major markets which had previously been handled by Venetian traders, during a period of high mortality in Venice. Furthermore, the Nobel Laureate Robert William Fogel considers it likely that the decrease of mortality in Britain in the 18th century paved the way for the industrial revolution. More specifically, Fogel suggests that the health improvements and better nutrition led to increases in labor force participation and productivity of workers. Finally, he calculates that these factors can explain around 30-50% of the economic growth of Britain in the years from 1790 to 1980.[1]

After the discovery of Penicillin by Alexander Fleming in 1928 many scientists began to believe that the battle against infectious diseases was coming to an end. For instance, Cockburn et al. (1963) predicted that within 100 years all major infections would have disappeared. Moreover, Russell et al. (1955) published an article on man's mastery of malaria. While it is true that some diseases like smallpox have been eradicated, other infectious diseases continue to exist and claim thousands of deaths every year. Moreover, malaria is far

[1] For further information see Fogel (1994), Fogel (2004) and Price-Smith (2001).

from being mastered and was still responsible for 2.2% of all deaths worldwide in 2002. Since the discovery of the human immunodeficiency virus (HIV) in the 80's, the view of being able to eradicate infectious diseases has slowly changed. The acquired immunodeficiency syndrome (AIDS) – caused by HIV – was a new infectious disease for which there is still no treatment available today. Besides AIDS, other new infectious diseases continue to appear. In the years from 1981 until 2006, 38 new germs were found (Turkington and Ashby 2007). Examples of newly emerging germs are HIV, avian flu, severe acute respiratory syndrome (SARS) and Creutzfeldt-Jakob disease (also known as human mad cow disease). Moreover, due to antimicrobial resistance new germs are an even more serious threat to health (Weber and Courvalin 2005).

Coming to the present day a study of the World Health Organization (2004) estimates that 57 million deaths occurred worldwide in 2002. In terms of death cause the most common causes of death were cardiovascular diseases and infectious diseases, with around one quarter of total deaths each.[2] Pinheiro et al. (2010) use a different method to compare the threat of different diseases. They calculate disability-adjusted life years, a measure also taking into account the non-lethal aspect of diseases.[3] In 2001, across all regions of the world, the average burden of disease amounted to 250 disability-adjusted life years per 1,000 inhabitants. Of these, 26% were due to infectious diseases which formed the primary cause, while cardiovascular diseases only account for 14%.[4] The vast majority of disability-adjusted life years due to infectious diseases (more than 98%) occur in low- and middle-income countries. However, for high income countries infectious diseases also represent a threat that should not be disregarded. Less than 100 years ago, the Spanish Flu killed between 50 and 100 Million people all over the world (Johnson and Mueller 2002). Public health experts warn that another pandemic can strike at any time (Turkington and Ashby 2007). The question is not if it will happen, but rather when (Barry 2005, p. 463). Moreover, it is estimated that a pandemic of a similar scale as the Spanish Flu would kill between 180 and 360 million people today (Osterholm 2005).

Historically, smallpox was the most severe disease in terms of total death toll. Smallpox already appeared in 1500 B.C. and was responsible for 200 million deaths in cumulative total (± 100 million due to low data availability the further one goes back in history) until its eradication in the 70's (see

[2]Other important causes were cancer and injuries with 12.5% and 9.1% respectively.

[3]This measure was developed in 1996 and is widely used by the World Bank and the World Health Organization. It is formed by the sum of years lost due to premature death and years lost due to disability (Murray et al. 2002).

[4]The main reason for this loss of importance of cardiovascular diseases is that infectious diseases kill many children and thus have a higher cost in terms of life years missed due to premature death.

for instance Hopkins 2002). Proceeding in chronological order, the Plague of Justinian killed as many as 25 million people (around 10% of the world population) troughout the world in the years in 541-542 (Kulikowski 2007). In the 14th century the Black Death killed an estimated 35 million in Europe – around 45-50% of Europe's population (Byrne 2006). The last two pandemics with similar death tolls appeared in the last century. After World War I, the Spanish Flu appeared with a total death toll of 50-100 million worldwide – around 5% of the world population (Johnson and Mueller 2002). Finally, AIDS was discovered in 1981 and no treatment is available until today. By 2007 a cumulative total of 24 million people had died from AIDS and by 2030 this cumulative total is predicted to reach 75 million (Bongaarts et al. 2010).

In this dissertation we analyze the economic consequences of infectious diseases. In particular, we study the effects of the Spanish Flu and HIV/AIDS on key economic indicators. What distinguishes these two conditions from other diseases is that most victims were prime-age individuals. Since they are at their peak in terms of productivity before their infections, the economic consequences should be even larger. Moreover, as we will show in the second part of this dissertation, there is also an important feedback from the economy which affects the spread of infectious diseases. More specifically, we study how the business cycle can enforce the spread of diseases causing procyclical sickness absence rates.

Health economics arose as a subdiscipline of economics in the 70's. It applies economic theory to health-related problems. Economic science was always concerned with the allocation of scarce resources. In the area of health it is important to find a price for the priceless (as described in the introduction to Health Economics, Zweifel et al. 2009). Such a price is needed to be able to compare the costs of different diseases and allocate limited resources in terms of finding cures for different health problems. Therefore, measuring the economic consequences of illnesses is a central topic in health economics. The estimates of the cost of illness are important for decision makers, since they offer an alternative perspective and can thus determine research priorities.

Moreover, one of the most fundamental questions in economics has always been the role of the state. Smith (1776, Book IV, Chapter IX, pg. 749) claims in the Wealth of Nations, that the first duty of the sovereign is protecting society. Even though Smith probably did not think about infectious diseases when he made this statement, one cannot deny that infectious diseases are a serious threat to society. Therefore, it seems important to study these diseases from an economic perspective.

Research in the field of infectious diseases by health economists only started quite recently. However, from an economic perspective this field seems fruitful.

Firstly, individual behavior in terms of prevention has large potential externalities, a topic already studied by economists for several decades. Secondly, given the threat infectious diseases pose to humans, studying their consequences in more detail and analyzing how the spread of diseases is affected by the economy appears important. Moreover, studying the economic outcomes can be helpful in establishing appropriate policy responses. The pandemics studied in this dissertation represent a large labor supply shock, which usually does not occur with such force. Thus, our results also are informative with respect to the empirical performance of macroeconomic models – since estimated outcomes can be compared to the diverging predictions of neoclassical and endogenous growth models.

In the first part of this dissertation we study the effects of HIV/AIDS and the Spanish Flu on different economic outcome variables. Estimating the relationship between these two diseases and economic outcomes is complicated by the fact that there are so many channels through which the epidemic might affect economic performance.

The first channel, affecting production factors immediately, involves *a)* a reduction in productivity of infected workers, and *b)* a possible reduction in average human capital due to replacement of skilled workers with less skilled workers. These channels might appear to be the most obvious ones, and yet, they have been contested in the literature. For example with respect to HIV, Bloom and Mahal (1997b) argue that these channels are of secondary importance since the most heavily affected countries have considerable reserves to draw upon. Nevertheless, there is evidence that in some countries agricultural production has been disrupted due to HIV (Gaffeo 2003). On the other hand, a reduced number of workers per unit of capital could in fact have a positive effect on per capita growth rates (Young 2007).

A second channel is the disruption to the transmission process of human capital that the pandemic imposes. For the Spanish Flu Boucekkine and Laffargue (2010) show that an increase in the number of orphans may have important distributional consequences in the long term. With respect to HIV/AIDS there are at least 16 million children aged 0–17 worldwide who have been orphaned due to HIV, and this figure has not yet started to decline (UNAIDS 2010). According to Bell et al. (2006), this channel might become very important in the long run, and may actually lead to an outright collapse in economic activity. The authors attribute these strong consequences to the fact that the disrupted human capital accumulation lingers for several generations, since orphans will also be unable to give their own children an appropriate education. A recent empirical study (Coneus and Mühlenweg 2011) estimates the impact of orphanhood on educational and health outcomes in eleven sub-Saharan

African countries. According to their estimates, there is strong evidence that orphans lag behind their non-orphaned peers in terms of educational outcomes. The estimated effect is 1/5 to 1/2 of a year. However, the group of 'social orphans', who live away from their biological parents although the latter are alive, consists of a much larger group and exhibits very similar patterns in terms of educational and health outcomes. It should further be noted that there will probably be lower investment in human capital in affected families even before orphanhood, to the extent that children have to leave school in order to care for their parents and carry out other household tasks (Gaffeo 2003).

In order to deal with the difficulty posed by various channels, we take different approaches. In a first step, we analyze the effect of HIV on GDP in African countries. Here we use a new econometric method developed by Abadie and Gardeazabal (2003), relying on synthetic control groups. For every country in consideration the method forms a weighted average of countries not affected by HIV. This average forms the counterfactual and thus allows estimation of the effects of HIV. The advantage of this method is that we do not have to model the single channels through which HIV affects the economy. However, this also makes it difficult to interpret the effects, because the method gives no information on what caused the effects.

In a second step we try to obtain a clearer understanding of how HIV affects GDP growth. Again we employ the synthetic control group method. However, instead of estimating the effects on GDP directly we look at demographic variables. This allows us to restrict the analysis to fewer channels.

Moreover, the analysis is complicated by the fact that different measures were taken in different countries in order to control the disease. This adds additional noise and makes estimation of effects difficult. Therefore, in a third step, we analyze the Spanish Flu in Sweden, a disease that occurred very rapidly and thus public health could not react. We analyze the effects in Sweden because of the fact that the disease spread shortly after World War I, in which Sweden was neutral. This is helpful because the World War could confound the effects.

In the second part of this dissertation we study the feedback effect of the economy on infectious diseases. The economy and its fluctuations pose different incentives which alter the behavior of humans. This is also true in terms of infectious diseases. As noted by McNeill (1977) migration of individuals was one of the most important factors allowing the spread of disease. Therefore, globalization increased the vulnerability of humans to infectious diseases due to a higher mobility and urbanization (Lederberg 1997).

Also, the business cycle was found to be an important determinant of

health. Ruhm (2000), for instance, finds that mortality varies with the business cycle, with more deaths during economic booms. The reason for this finding is revealed when mortality is divided by its cause. Ruhm (2000) notes that the main reason are more traffic accidents during booms leading to higher mortality. For sickness absence, the literature also suggests a procyclical relationship. Here the explanations are based on labor force composition and reduced moral hazard due to fear of unemployment. However, so far the cause for sickness absence has not been taken into consideration. We decompose sickness absence by causes and find that the procyclical pattern appears only for infectious diseases. Therefore we propose a new explanation, where the business cycle alters the spread of infectious diseases leading to higher sickness absence during economic booms.

Investigating this relationship is important since worker absenteeism is costly for firms, especially during times of upswing of the economy, where labor demand is higher (see for instance Audas and Goddard 2001). Identifying the cause therefore helps to reduce costly absenteeism.

This dissertation is structured as follows. Chapter 2 analyzes how HIV prevalence affects the Gross Domestic Product (GDP) in the twelve most heavily affected countries, i.e. the countries which have ever exhibited a HIV prevalence rate of 10% or more. As we show in Chapter 2, the effects of HIV on GDP growth are very heterogeneous. This is probably due the high number of channels through with HIV affects the economy. In Chapter 3 we analyze demographic effects of HIV in order to restrict the possible channels and gain more insights about the channels generating the effect heterogeneity. More specifically, in Chapter 3 we analyze how HIV affects life expectancy, and death and birth rates. Again some effect heterogeneity with respect to the fertility rate persists. Another reason for the effect heterogeneity is that different countries took different measures to prevent the spread of the disease. Therefore, we provide an in-depth analysis of one country in Chapter 4. In Chapter 4 we analyze the economic consequences of the Spanish Flu, a disease that occurred very rapidly and thus prevention was very limited. In Chapter 5 the causal channel is reversed and the relationship between business cycles and the spread of diseases is analyzed. This Chapter provides a new explanation for the procyclical nature of sickness absence caused by presenteeism (i.e. working while sick) and infections. Finally, Chapter 6 concludes.

CHAPTER

2

The Economic Impact of the HIV Pandemic [‡]

2.1 Motivation

The main purpose of this Chapter is to analyze the impact of the HIV pandemic on economic performance. In particular, it is our goal to apply a new method and a new identification strategy in order to estimate the causal effect of the pandemic in some particularly strongly affected countries. Our method relies on a careful selection of suitable comparison countries, and allows the effect of HIV on economic performance to vary not only with the prevalence rate, but also between countries and time periods. Moreover, we employ some new high-quality datasets which improve our ability to perform an empirical analysis.

The previous literature has typically estimated versions of an augmented Solow model using dynamic panel data techniques. Starting from initial estimates using simple cross-sectional techniques, the methods used have become increasingly sophisticated: after the difference GMM estimator (Arellano and Bond 1991), the systems GMM estimator has been introduced (Blundell and Bond 1998, Maclaine 2006). Some authors also explicitly take the possible endogeneity of the AIDS epidemic into account: McDonald and Roberts (2006) estimate a separate reduced form equation for health. Cahu and Fall (2011) use a similar approach but rely on a dynamic prediction of the prevalence

[‡]This Chapter is based on Karlsson and Pichler (2012b).

rates. Even if the evidence supporting a reverse causality running from economic performance to HIV incidence is weak (Durevall and Lindskog 2009), other endogeneity problems related to omitted variables still abound. Moreover, Goenka and Liu (2012) show in a recent paper that economic effects of diseases are highly non-linear. Therefore estimating a reduced model where some economic outcome variable depends on the prevalence rate will yield biased estimators.

Leaving endogeneity aside, the empirical specification of the relationship between AIDS and economic growth is complicated by the fact that there are so many channels through which the epidemic might affect economic performance. Some channels which are similar for HIV and the Spanish Flu and have already been analyzed in the introduction. There we mentioned that these diseases decrease productivity and might affect human capital. Moreover, since both HIV and the Spanish Flu kill prime-age individuals dependents might be left behind, and in the case of orphans they might fail to acquire skills normally taught to them by their parents. On top of these channels already discussed, the specificity of HIV gives rise to additional channels which might harm the economy. In particular, the alteration of *individual incentives* and effects due to changing *relative prices* should also be considered.

The AIDS epidemic is assumed to have an impact on growth through changing individual incentives for people living in affected areas. An increased demand for health care might lead to reduced savings rates. Moreover, the reduction in life expectancy induced by the disease is likely to lead to weaker incentives to save and to invest in human capital, and thus reduce growth (Boucekkine et al. 2002, Fortson 2011).

Fortson (2011) analyses the impact of the regional AIDS prevalence rate on human capital investment. She finds that a 10 per cent increase in the prevalence rate is associated with a reduction in average years of schooling by half a year. Considering the low level of schooling in affected countries, this is a large effect, and it cannot be attributed to direct effects like orphanhood or care for infected family members. Instead, the empirical evidence is suggestive of reduced longevity affecting the optimal human capital investment. On the other side of the labor market, an increased turnover of workers may reduce incentives of employers to invest in skills and training. Finally, the incentives to engage in sexual activity will undoubtedly change, with possible implications for fertility and thus growth. However, there is remarkable disagreement as to the effects of HIV on fertility (Durevall and Lindskog 2011). The reason may be that there are several different mechanisms at work – preference for fertility, avoidance of infection, longevity-induced effects – which work in different directions. Young (2005, 2007) identifies a large negative effect of

the epidemic on fertility, whereas other studies (Fortson 2009, Kalemli-Ozcan 2006, Juhn et al. 2008) report a much smaller or even positive effect. According to Durevall and Lindskog (2011), the effect appears to be heterogeneous and depend on the age and the number of previous children.

Moreover, as relative prices change in response to the pandemic, behaviors of investors may be affected as well. The returns on investment in affected countries might change for two reasons. Firstly, the pandemic imposes costs on the private sector which are in some parts similar to the costs of the public sector mentioned above (Liu et al. 2004). Secondly, the capital deepening that possibly occurs due to the illness will reduce returns on capital. To the extent that these changes discourage foreign direct investment, there will be a second-order effect on growth.

Concerning the estimated *effects*, the comparability of studies is limited to some extent, since some of them focus on the marginal effect of a small increase in the prevalence rate, whereas others estimate the overall effect towards a benchmark scenario of no AIDS at all. Some earlier studies on heavily affected countries estimated an overall reduction in the growth rate of 1.2-1.7 per cent (Gaffeo 2003) – which obviously translates in to considerable long-run effects. More recent studies, using more sophisticated techniques, reach somewhat less dramatic conclusions. For example, Cahu and Fall (2011) estimate that GDP per capita will be 12 % lower in the long term compared to a no-AIDS scenario.

In our analysis we do not need to restrict ourselves to one channel. We are able to see the total effect of these channels on GDP per capita. Our results suggest that there is substantial heterogeneity in the effects. For some countries, we fail to identify an effect at the per capita level as well. However, there are three countries in the dataset which have suffered a large impact of HIV on their living standards. For these countries, the estimated effect ranges from from -25 to -77% in terms of GDP per capita, or -1.5 to -5.5% in terms of annual growth rate.

In a final step we employ regression analysis to look at the overall relationship between GDP and deviation from a previous growth path. Here we estimate that on average one percentage point of HIV prevalence costs 3-4% in terms of GDP. Furthermore, in this anaylsis we can rule out that our results are driven by an "Africa effect" often found in the empirical growth literature. Due to the fact that we are able to find a general relationship, but we still observe a high level of heterogeneity between countries, this Chapter complements case studies that analyze how HIV disrupted the economcy within a certain country.[5]

This Chapter is structured as follows: Section 2.2 gives a brief introduction

[5]See for instance Arndt (2006) for a case study in Mozambique,(MacFarlan and Sgherri 2006) for Botswana and Robalino et al. (2002) for Kenya.

on the HIV pandemic. In Section 2.3 we present the econometric approach, while Section 2.4 will describe the data. We will first focus the most strongly affected countries and estimate a treatment effect for each country individually in Section 2.5. In Section 2.6 we will look at all affected countries and analyze heterogeneity in estimated treatment effects. Finally, we discuss the obtained results in Section 2.7.

2.2 The AIDS pandemic: some stylized facts

In this Section, we give a brief overview of the evolution of the AIDS pandemic from its very beginning up until the most data available, with a particular focus on aspects of the pandemic that are of importance to the current study.

The acquired immune deficiency syndrome (AIDS) was first recognized in 1981. According to the latest estimates from the UNAIDS agency, the worldwide prevalence was 33.3 million as of 2009 (UNAIDS 2010). Although the prevalence rate has been increasing over the past decade, there are some signs that the growth rate of pandemic decreased and there are clear signs of recovery in some heavily affected countries. According to the latest available figures, incidence has dropped by around a fifth during the last decade (UNAIDS 2010).

However, the prevalence of the disease varies significantly between countries and also between regions within countries. By all measures, Sub-Saharan Africa is by far the most heavily affected region in the world: 22.5 million, or two-thirds of the world's total AIDS-infected population resides in a Sub-Saharan country. Some further comparisons between world regions are provided in Table 2.1.

The current HIV prevalence rate in Sub-Saharan Africa is estimated at 7.2 per cent. Women are generally more strongly affected than men, with the proportion of HIV infected individuals being 56 per cent female.

2.2.1 Evolution of the Pandemic: Affected Countries

We now take a closer look at the evolution of the AIDS pandemic in Sub-Saharan Africa since the 1970s. In this part, we use data from Oster (2007), which have been derived based on UNAIDS statistics, using earlier information on trends form UNAIDS and linear interpolation. The UNAIDS statistics are in turn based on population-based testing, antenatal clinic data, expert advice from country-level HIV organizations, and epidemic modeling. There is a severe lack of reliable information to go by, but the dataset we use appears to be the best available. Oster (2010) considers an alternative method which bases prevalence estimates on age-specific HIV-related mortality and information on

Table 2.1: HIV Statistics by Region

Region	**Year**	**Adults & Children Living with HIV**	**Adults & Children Newly Infected**	**Adult prev. rate (15–49)**	**AIDS related Deaths (Adults)**
Sub-Saharan Africa	2009	22.5 M [20.9–24.2 M]	1.8 M [1.6–2.0 M]	5 [4.7–5.2]	1.3 M [1.1–1.5 M]
	2001	20.3 M [18.9–21.7 M]	2.2 M [1.9–2.4 M]	5.9 [5.6–6.1]	1.4 M [1.2–1.6 M]
Middle East & North Africa	2009	460 000 [400,000–530,000]	75 000 [61,000–92,000]	0.2 [0.2–0.3]	24 000 [20,000–27,000]
	2001	180 000 [150,000–210,000]	36 000 [32,000–42,000]	0.1 [0.1–0.1]	8300 [6,300–11,000]
South and South-East Asia	2009	4.1 M [3.7–4.6 M]	270,000 [240,000–320,000]	0.3 [0.3–0.3]	260,000 [230,000–300,000]
	2001	3.8 M [3.5–4.2 M]	380,000 [350,000–430,000]	0.4 [0.3–0.4]	230,000 [210,000–280,000]
East Asia	2009	770,000 [560,000–1.0 M]	82,000 [48,000–140,000]	0.1 [0.1–0.1]	36,000 [25,000–50,000]
	2001	350,000 [250,000–480,000]	64,000 [47,000–88,000]	<0.1 [< 0.1–< 0.1]	15,000 [9,400–28,000]
Oceania	2009	57,000 [50,000–64,000]	4,500 [3,400–6,000]	0.3 [0.2–0.3]	1,400 [<1,000–2,400]
	2001	29,000 [23,000–35,000]	4,700 [3,800–5,600]	0.2 [0.1–0.2]	<1,000 [<500–1,100]
Central and South America	2009	1.4 M [1.2–1.6 M]	92,000 [70,000–120,000]	0.5 [0.4–0.6]	58,000 [43,000–70,000]
	2001	1.1 M [1.0–1.3 M]	99,000 [85,000–120,000]	0.5 [0.4–0.5]	53,000 [44,000–65,000]
Caribbean	2009	240,000 [220,000–270,000]	17,000 [13,000–21,000]	1 [0.9–1.1]	12,000 [8,500–15,000]
	2001	240,000 [210,000–270 000]	20,000 [17,000–23,000]	1.1 [1.0–1.2]	19,000 [16,000–23,000]
Eastern Europe and Central Asia	2009	1.4 M [1.3–1.6 M]	130,000 [110,000–160,000]	0.8 [0.7–0.9]	76,000 [60,000–95,000]
	2001	760,000 [670,000–890,000]	240,000 [210,000–300,000]	0.4 [0.4–0.5]	18,000 [14,000–23,000]
Western and Central Europe	2009	820,000 [720,000–910,000]	31,000 [23,000–40,000]	0.2 [0.2–0.2]	8,500 [6,800–19,000]
	2001	630,000 [570,000–700,000]	31,000 [27,000–35,000]	0.2 [0.2–0.2]	7,300 [5,700–11,000]
North America	2009	1.5 M [1.2–2.0 M]	70 000 [44,000–130,000]	0.5 [0.4–0.7]	26 000 [22,000–44,000]
	2001	1.2 M [960,000–1.4 M]	66,000 [54,000–81,000]	0.4 [0.4–0.5]	30,000 [26,000–35,000]
TOTAL	2009	33.3 M [31.4–35.3 M]	2.6 M [2.3–2.8 M]	0.8 [0.7–0.8]	1.8 M [1.6–2.1 M]
	2001	28.6 M [27.1–30.3 M]	3.1 M [2.9–3.4 M]	0.8 [0.7–0.8]	1.8 M [1.6–2.0 M]

Source: UNAIDS (2010)

life expectancy after infection. The prevalence rates that emerge from that analysis follow dynamics similar to those found in the UNAIDS data, but for some countries the two measures diverge. This does not represent a major challenge to our identification strategy, but it would of course be useful to consider the sensitivity of estimates with respect to the prevalence rates.

For our analysis, it is useful to establish two general thresholds for the prevalence of HIV.[6] Thus, from now on we will refer to a country as *strongly affected* if the estimated prevalence rate is above 1 per cent.[7] Likewise, we refer to countries with prevalence rates above 10 per cent as *severely affected*. The rest of the countries, i.e., those that have never crossed the 1 per cent threshold, will be referred to as *mildly affected*. This categorization of countries may seem arbitrary, however it allows us to go from a continuous treatment to a discrete one. It gives us twelve countries to analyze more in detail and compare the results across different countries. Later, in Section 6, we will look at all strongly affected countries.

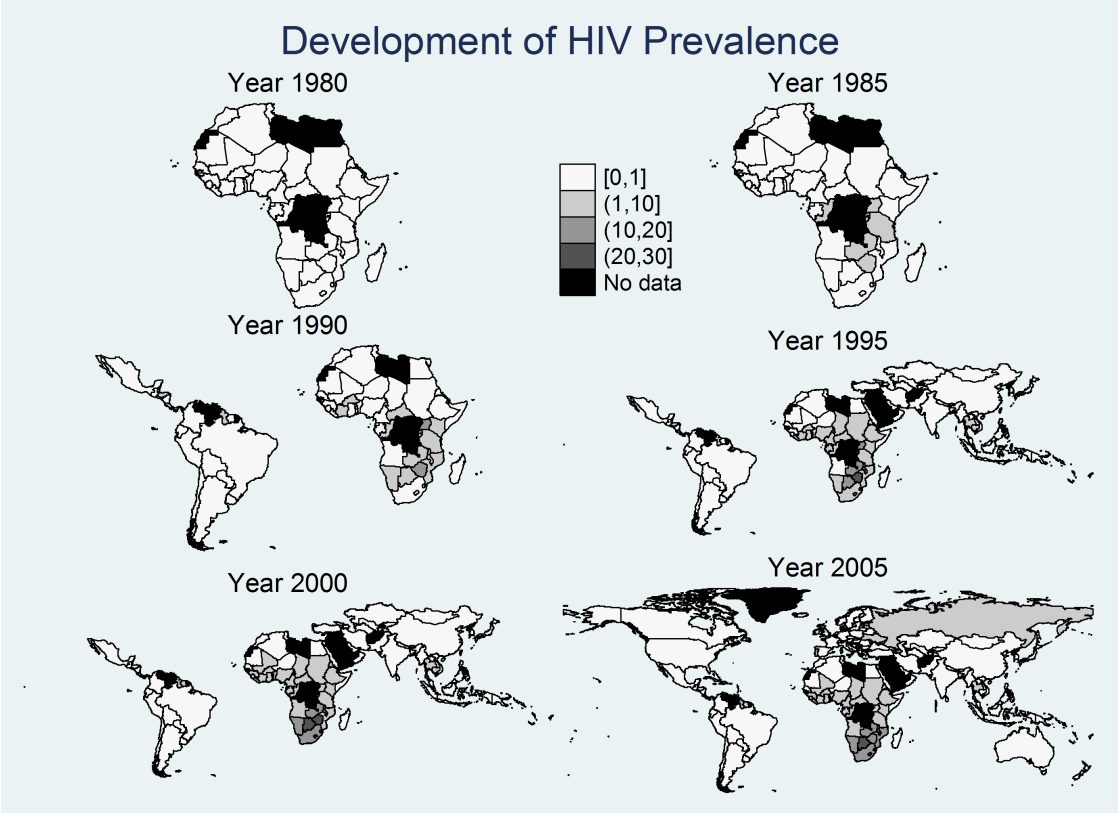
Our method allows for some degree of endogeneity in the treatment indicator – in the sense that the occurrence of the intervention may be correlated with unobservables. Nevertheless, it seems reasonable to assume that in the early stages, the spread of the disease was relatively exogenous. As Figure 2.1 shows the spread of the disease appears to be mainly driven by geographical factors (all countries in the Sub-Saharan region are hit by the disease), while the actual prevalence level reached in each country will also be influenced by other factors, like poverty for example. Thus, treatment assignment is reasonably exogenous, whereas the development and wider spread of the disease is probably not. Summing up we will contrast severely affected countries with mildly affected countries, what matters is that the 'treatment group' (severely affected countries) and the 'control group' (mildly affected countries) differ in a way that is interesting for policy purposes.

Available evidence suggests that the pandemic spread at a very high speed during the early years in affected countries. For example, in Malawi, the first AIDS case was diagnosed in 1985, a 1 % prevalence rate was reached in 1988, and in 1990 the prevalence rate was at 2 % already (Arrehag et al. 2006). Moreover, our data shows that the average time it took from a prevalence rate of 0.5% to 1% was 1.1 years, while only 14 out of 91 mildly affected countries ever reached a prevalence rate greater or equal to 0.5%. Finally, given that the incubation period is some 6-8 years on average, this means that governments and aid agencies had very small possibilities to moderate the epidemic during the early years (Gaffeo 2003).

[6] We only have data on HIV prevalence, therefore what we will estimate is the empirical relationship between HIV and growth (and not AIDS directly).

[7] Bell et al. (2006) consider a similar threshold to determine the start of the pandemic.

Figure 2.1: Spread of HIV



Some summary statistics are provided in Table 2.2. In the first column, we provide the estimated HIV prevalence rate as of 2007. The next column shows when a country first crossed the threshold of 1 per cent prevalence. Clearly, all countries in the sample but Niger have done so to date, but the year of crossing varies from 1982 (Uganda) to 2007 (Senegal). Most countries entered the strongly affected group in the late eighties or the early nineties.

The last column in Table 2.2 reports the year when the countries are estimated to have crossed the 10 per cent threshold to become *severely* affected (if ever). Twelve countries in the sample have done so, and the estimated timing varies between 1987 and 2001. Interestingly, some Eastern African countries (Kenya, Rwanda, Uganda) that observed prevalence rates above ten per cent have afterwards reduced their rates to a lower level. The observation that the severely affected countries diverge over time is also confirmed in Figure 2.2, which plots prevalence rates over time.

2.2.2 Control Group

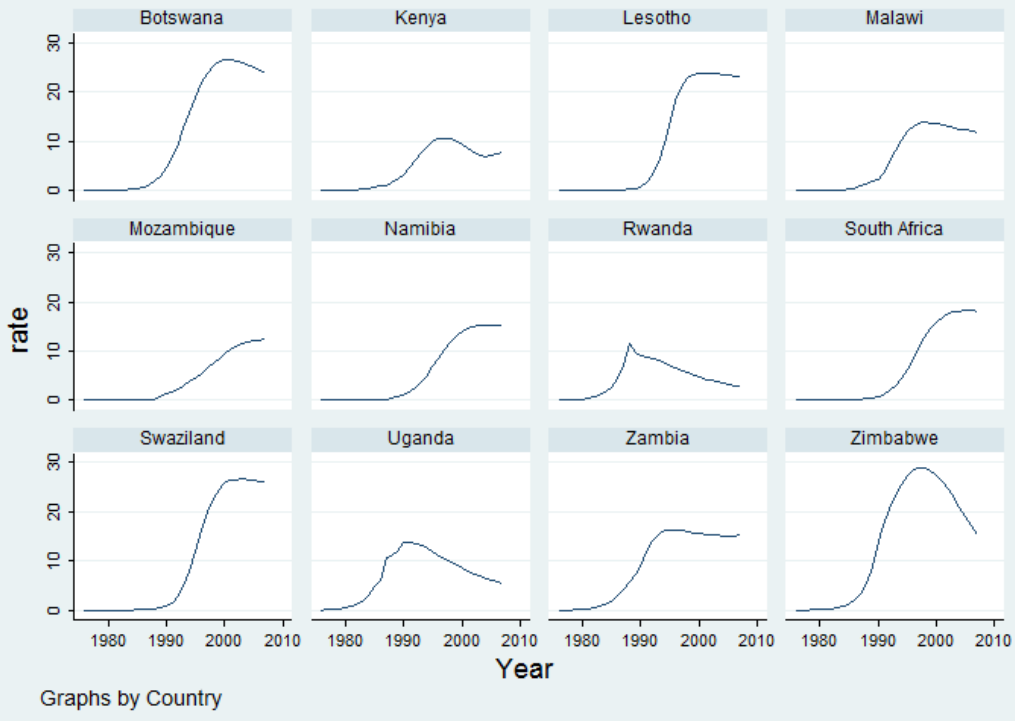
We defined our control group as countries which, according to UNAIDS estimates, have never crossed the 1 per cent prevalence threshold, i.e. mildly affected countries according to the terminology above. As Table 2.3 reveals, there is a large number of countries available which are at different levels of

Table 2.2: HIV Prevalence Rates

Country	**Rate**	**Started**	**Reached 10+**
Angola	2.1	1997	.
Benin	1.2	1996	.
Botswana	23.9	1987	1993
Burkina Faso	1.6	1987	.
Burundi	2	1985	.
Cameroon	5.1	1988	.
Central African	6.3	1986	.
Chad	3.5	1988	.
Congo	3.5	1985	.
Cote d'Ivoire	3.9	1987	.
Equatorial Guine	3.4	1990	.
Eritrea	1.3	1996	.
Ethiopia	2.1	1991	.
Gabon	5.9	1991	.
Ghana	1.9	1994	.
Guinea	1.6	1999	.
Guinea-Bissau	1.8	1996	.
Kenya	7.8	1987	1996
Lesotho	23.2	1991	1995
Liberia	1.7	1993	.
Malawi	11.9	1988	1994
Mali	1.5	1997	.
Mozambique	12.5	1990	2001
Namibia	15.3	1990	1997
Niger	.8	.	.
Nigeria	3.1	1991	.
Rwanda	2.8	1983	1988
Senegal	1	2007	.
Sierra Leone	1.7	1995	.
South Africa	18.1	1991	1997
Swaziland	26.1	1991	1995
Togo	3.3	1991	.
Uganda	5.4	1982	1987
United Republic	6.2	1985	.
Zambia	15.2	1984	1991
Zimbabwe	15.3	1985	1990

Source: Own calculation based on Oster (2007)

Figure 2.2: Evolution of HIV Prevalence Rates in Severely Affected Countries



development.

There are forty additional countries for which HIV prevalence statistics are available, but since these belong to the strongly but not severely affected countries, we could not use them in our main analysis. The vast majority of these countries are African, but notable exceptions are Ukraine, Jamaica and Thailand.

2.3 Econometric Approach: Synthetic Control Groups

Traditionally, comparative case studies of this kind rely on comparisons between affected and non-affected units, using aggregate data. This type of analysis differs from standard microeconomic evaluation studies in two important respects. Firstly, the choice of comparison units is not as straightforward as for micro studies. Secondly, since data tend to be collected at the population level, sampling error is not an issue. Hence, standard tools for inference do not apply.

The synthetic control group approach (Abadie et al. 2010) has been developed to address these issues,

- providing a transparent, data-driven method for selection of control units,

Table 2.3: Control Group: Mildly Affected Countries

Country	**Prev 2007**	**Country**	**Prev 2007**
Algeria	.1	Argentina	.4
Armenia	.1	Australia	.1
Austria	.3	Azerbaijan	.1
Bangladesh	.1	Belarus	.2
Belgium	.2	Bhutan	.1
Bolivia	.2	Bulgaria	.1
Cambodia	.6	Canada	.2
Chile	.4	Colombia	.6
Comoros	.1	Costa Rica	.3
Croatia	.1	Cuba	.1
Czech Republic	.1	Denmark	.2
Dominican Republ	.8	Ecuador	.4
Egypt	.1	El Salvador	.8
Fiji	.1	Finland	.1
France	.4	Georgia	.1
Germany	.1	Greece	.1
Guatemala	.7	Hungary	.1
Iceland	.3	India	.4
Indonesia	.1	Iran	.2
Ireland	.2	Israel	.2
Italy	.3	Japan	.1
Kazakhstan	.1	Korea Rep	.1
Kyrgyzstan	.2	Lao PDR	.2
Latvia	.6	Lebanon	.1
Lithuania	.1	Luxembourg	.3
Madagascar	.2	Malaysia	.5
Maldives	.1	Malta	.1
Mauritania	.7	Mexico	.3
Mongolia	.1	Morocco	.1
Myanmar	.6	Nepal	.4
Netherlands	.2	New Zealand	.1
Nicaragua	.2	Norway	.1
Oman	.1	Pakistan	.1
Papua New Guinea	.9	Paraguay	.3
Peru	.4	Philippines	.1
Poland	.1	Portugal	.5
Qatar	.1	Republic of Mold	.4
Romania	.1	Senegal	.8
Serbia	.1	Singapore	.1
Slovakia	.1	Slovenia	.1
Somalia	.6	Spain	.4
Sri Lanka	.1	Sweden	.1
Switzerland	.4	Tajikistan	.2
Tunisia	.1	Turkey	.1
United Kingdom	.2	United States of	.6
Uruguay	.5	Uzbekistan	.1
Vietnam	.4		

Source: Own calculations based on UNAIDS data.

which is not based on extrapolation.

- offering alternative tools for inference in aggregated data analysis.

An additional strength of this method is that no information on post-intervention outcomes is needed to design the study. Thus, there is a much lower risk that the study design is biased. Moreover, non-linearities in the outcome variable are not a problem since the method will only employ a few thresholds in terms of prevalence, rather than prevalence rates at each point in time. Abadie and Gardeazabal (2003) used this method for estimating the effect of terrorism on GDP per capita in northern Spain and the effect of a Californian anti-Tobacco law in 1988 on cigarette consumption (Abadie et al. 2010). Examples of other studies using this method are Hinrichs (2011) who studies the effects of affirmative action bans on minority students, or Montalvo (2011) studying the effects of terrorism on voting behavior.

2.3.1 Assumptions

We now present the main assumptions needed for the analysis. Since the HIV epidemic reached critical levels in different countries at different points in time, each affected country needs to be analyzed separately. Thus, in our notation below we proceed as if there were only one single treated unit and J further control units.

We denote by Y_{it}^N the natural logarithm of GDP per capita that would have been observed in country i at time t in absence of the HIV epidemic. Also, we denote by T_0 the last pre-intervention period, i.e.

$$T_0 = \inf\{t | R_{1t} \geq 0.01\} - 1$$

where R_{1t} denotes the HIV prevalence rate in the affected country.

Moreover, let Y_{it}^I be the GDP per capita that would be observed for unit i at time t if the epidemic had started to take off at time $T_0 + 1$. Since HIV is unlikely to have had an effect on GDP before the outbreak of the pandemic, we also have $Y_{it}^I = Y_{it}^N \forall t = 1, \ldots, T_0, \forall i = 1, \ldots, J + 1$.

A further assumption which is needed is that the outcomes of non-treated countries is unaffected by the intervention. There are several ways in which this assumption could be violated. Firstly, there is obviously the risk of contagion, to the extent that the countries of the control group also become affected by the pandemic. For example, Oster (2007) delivers strong evidence that trade between countries elevates incidence rates. However, this possibility has been eliminated by considering only countries which have had very low prevalence rates throughout. Secondly, the outburst of the HIV epidemic could

affect countries in the control group via *migration*, *aid flows*, *foreign direct investment*, and *international trade*.

Concerning *migration*, the outflows from Sub-Saharan Africa are likely to be too small to have had an impact on the receiving countries' economies. For example, the total stock of immigrants from Sub-Saharan countries in OECD countries was around 3.9 million in 2002, working out at less than 0.5 per cent of the total population (OECD 2005). The possibility of HIV-induced migration between Sub-Saharan countries cannot be entirely discarded, but it is difficult to see how this could possibly have a discernible impact on aggregate economic performance (United Nations 2009).

Concerning financial flows between countries – aid, FDI and trade – there are two possible effects to consider. One is that the HIV epidemic in the treated units has a direct effect on the developed countries from which these flows typically originate. This is again unlikely, considering the tiny shares of these countries' economies that these flows make out. Besides, these rich countries are unlikely to be important as controls anyway. A more serious threat to the identification strategy is the possibility that other poor countries are affected: international and bilateral aid may be diverted to Sub-Saharan countries due to the HIV epidemic, and FDI may be diverted away for the same reason. Even though we are well aware of this problem, we think it can be disregarded for two reasons. Firstly, with 865 million inhabitants as of 2010, Sub-Saharan Africa comprises only 15 % of the world's total LDC population (Haub 2010). Thus, to the extent that FDI and international trade have been disrupted due to the HIV epidemic and diverted to other countries, the effects have probably been quite thinly spread. Secondly, considering the ongoing debate on aid effectiveness (Banerjee et al. 2007), it is unlikely that an HIV-induced diversion of aid to Sub-Saharan countries has led to discernible negative effects in the countries that might otherwise have received the aid.

Next, define $A_{it} \equiv Y_{it}^I - Y_{it}^N$ as the effect of the HIV epidemic for unit i at time t, and let D_{it} be an indicator which takes on the value one whenever the HIV epidemic has crossed the 1 per cent threshold: $D_{1t} = \mathbf{1}(t > T_0)$. Thus, the observed outcome for unit i at time t equals

$$Y_{it} = Y_{it}^N + A_{it} D_{it} \tag{2.1}$$

2.3.2 Modeling Growth

The synthetic control method does not require an explicit theoretical model for the outcome variable, but given the rich theoretical and empirical literature on economic growth and development, it appears reasonable to frame the

estimation procedure with reference to growth theory. Abadie et al. (2010) suggest the following generic equation for the outcome in the non-intervention case:

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \tag{2.2}$$

where δ_t is a time effect common to all units, θ_t is a vector of possibly time-dependent coefficients, λ_t is a vector of unobserved common factors, and μ_i is a vector of unknown factor loadings. Clearly, we may think of Equation 2.2 as a growth regression in reduced form, where the autoregressive element has been eliminated by insertion.[8] Thus, initial value of GDP, Y_{i0} should clearly be included amongst the regressors Z_i. Furthermore, we include gross capital formation, indicators of the labor force, sectoral composition, human capital, population growth, rural and total population.

Moreover, there is a large empirical literature that is very useful to identify variables which are important determinants of economic growth and development. Much of this literature actually revolves around the so-called "Africa dummy", i.e., whether the poor growth performance of Sub-Saharan Africa is attributable to observable differences or not. The starting point of this literature was a study by Barro (1991), according to which there was a negative growth effect associated with being a country in Sub-Saharan Africa. A paper by Sachs and Warner (1997) also considered Sub-Saharan Africa's poor growth record during the period 1965-1990. Their main finding was that the growth experience of African countries can be studied within the same framework as for other countries; in other words, there is no specific effect associated with being a country in Sub-Saharan Africa. Factors found to be particularly important determinants of growth were, in diminishing order of importance, openness to trade; life expectancy; institutional quality; central government saving; natural resources; relative growth of the labor force; climate; and whether the country was landlocked or not. Literature reviews by Collier and Gunning (1999a,b) from the same time generally confirm these findings; however, they challenge the view that the landlockedness or population growth rates may be seen as exogenous. A good review of the entire growth regression literature, with a particular focus on Africa, is provided by Ndulu et al. (2008).

Several authors have tried to explain the poor macroeconomic performance in Africa with reference to their poor institutions and their colonial past. A seminal paper by Acemoglu et al. (2001) found that institutions matter a lot more than what a simple cross-sectional regression reveals; however, that paper has since been heavily criticized (Albouy 2008). Nevertheless, there is

[8] Abadie et al. (2010) show that the synthetic control approach is unbiased also in an autoregressive model

widespread consensus that institutions matter. Englebert (2000) introduces the concept of "state legitimacy", which is generally – but not uniformly – worse in Sub-Saharan Africa than in other parts of the world. He finds that the colonial inheritance of poor state legitimacy explains not only inferior growth, but also to a large extent policies that are harmful to growth. Also Bertocchi and Canova (2002) attribute a large portion of sluggish growth in Africa to its colonial inheritance. Even though these studies are very useful to generate hypotheses concerning the variables that are associated with economics growth in an African context, it should also be noted that the growth regression literature has been criticized. In a recent contribution, Jerven (2009) challenges the view implicit in many growth regressions, that the African performance has been uniformly worse since independence. In fact, the period of sluggish growth – predominantly from the mid-1970s to the mid-1990s – was preceded by a period of higher-than-average growth on the African continent. This is an important point; however, our data-driven empirical strategy has been designed precisely to allow for that kind of irregularities.

2.3.3 Implementation

The synthetic control method involves estimating two matrices: $\mathbf{V}$ is the weighting matrix determining the relative predictive power of various outcome variables Z_i and of the outcome variable itself. The vector $\mathbf{W}$ is a vector of non-negative weights given to the J control countries.

The criterion minimized is given by

$$\left\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0\mathbf{W}\right\|_{\mathbf{V}} = \sqrt{\left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0\mathbf{W}\right)' \mathbf{V} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0\mathbf{W}\right)} \tag{2.3}$$

where $\bar{\mathbf{X}}_j$ is a vector of averages over the pretreatment period of elements of Z_i and Y_i, for treated and control units, respectively. This will give us an optimal country weight matrix among all diagonal positive definite matrices depending on the variable weight ($W^*(V)$).

The predictors V are chosen such that the average distance from of the outcome variable, i.e. the root of the mean squared prediction error is minimized in pre-intervention periods:

$$RMSPE = \sqrt{\frac{\sum_t \left(Y_{1,t} - \mathbf{Y_{0,t}}\mathbf{W}^*(\mathbf{V})\right)^2}{N}} \tag{2.4}$$

where N represents the number of years in the pretreatment period.

2.4 Data and Variables

For our analysis we use data from various sources. As already mentioned above the main source of our HIV data is Oster (2007). Our dependent variable is GDP per capita at purchasing power parity (PPP).[9] In order to enrich this dataset we also used population data from Rosling (2012) and GDP data from the Penn World Tables (Alan Heston et al. 2006). All variables used and their sources are presented in Appendix A. As mentioned earlier, we use different predictor variables in order to find a close match of the treated country as synthetic control unit.

Combining these datasets, we are able to construct a balanced panel dataset for almost all countries from 1962 to 2007 with only a few missing values for some countries. The median year of the intervention (when the HIV prevalence rate increases to more than 1%) is 1987, with minimum of 1982 and maximum of 1991, which gives us on average 25 pre-intervention years and 22 post-intervention years. [10]

As discussed in Section 2.3.3, the choice of variables included in $\mathbf{X}$ will influence the countries the estimator uses to replicate the treated country and will thus also affect the estimates. We follow two complementary strategies in order to obtain a reliable estimate: in a first set of estimates, we rely on the theoretical literature on economic growth and include variables motivated by standard economic models, while in a separate set of estimates, we base our estimation on the empirical growth literature for the choice of our variables in $\mathbf{X}$.

We embed the first step of our analysis in an augmented Solow growth model. Our predictors for GDP are as follows: gross capital formation; population in the labor force; total population; population growth; proportion of population living in rural areas and the value added of the agricultural, industrial and service sector respectively.[11] Moreover, we include human capital from two sources: years of schooling as provided by Barro (2001) and proportion of the population who completed primary, secondary and tertiary education respectively, provided by Lutz et al. (2007) and transformed into

[9] We also performed a similar analysis using total GDP as dependent variable. However, as Oster (2007) finds the median time to death from infection to death is around ten years, also in African countries. Therefore it is not too surprising that we do not find any effects in the aggregate. As Table 2.2 shows most of our treated countries hit the prevalence rate of 10% in the 90's. Consequently the effects of mortality should only be seen 10-15 years later and so these effects can be observed only recently or even in the future.

[10] An issue in this context is that many countries are not independent at the beginning of our analysis. Data is available also for them, even though the influence of parent states is not clear. Therefore, as a robustness check, we drop all pre-independence observations. The results however were not affected by this change.

[11] The source of these variables and the GDP variable is the World Bank database (Mundial 2011)

years of schooling by us.

In a second set of estimates we relate our estimation approach to the empirical literature mentioned in Section 2.3.2. In these empirical studies it was shown that many different variables may play an important role for economic growth. Moreover, as shown in equation (2.3) the growth indicators are averaged over the pre-intervention period. One could imagine that information on the colonial origin of a particular country or landlockedness might have a higher predictive power for long term GDP growth than the capital stock averaged over 20 years. Therefore in a second step we include different variables often used in empirical growth studies and employ a data-driven strategy, such that the vector of variables $\mathbf{X}$ is allowed to differ from country to country. In these data-driven specifications, the variables are chosen in a way to obtain the best match between the treated unit and its synthetic control group in the pre-intervention period. This way we find variables that are closely related to the growth trajectory in each country.

These additional data come from various empirical studies. We use the data from (Mundial 2011) on percent of females in labor force and arable land and total land area in hectares. The climate and soil suitability of a country determine its potential for food production and thus the possibility to meet the basic needs of the population. For this reason we included several agricultural suitability variables, such as climate and soil quality and data on physical geography (e.g. average meters above sea level) and population from Gallup et al. (2001). Furthermore, we obtained data on ethnic and cultural fractionalization from Fearon (2003) and here the logic is that a country with a very high degree of fractionalization might have more frequent conflicts disrupting economic growth. Moreover, we merged our dataset with the dataset from Ciccone and Jarocinski (2010) which includes several determinants of economic growth, such as religion, influences from colonialization, life expectancy, fertility and many more. Many economic studies indicate that institutions might play an important role (see, e.g. Rodrik and Wacziarg 2005). In order to take them into account we included several variables from La Porta et al. (1999) and Englebert (2002) measuring the quality of the government along several dimensions such as bureaucratic delays, the estimated size of the black market, the level of corruption and other variables. We also include regime characteristics, using the Polity IV authority variable (Marshall et al. 2009), which measures the political authority on a scale ranging from -10 ('hereditary monarchy') to +10 ('consolidated democracy'). Moreover, Abadie and Gardeazabal (2003) show that violence and terrorism significantly affected economic outcomes in the Basque Country. This may well be the case in other countries as well, and therefore we also control for political violence within a country using the

Major Episodes of Political Violence dataset (Marshall 2010), which measures violence within a country on a zero (peace) to ten (highest violence level) scale. Finally, the globalization index provided by Dreher (2006) was used to measure overall globalization. Some empirical papers found that the ratio of GDP to GNP, a proxy for structural openness (see for instance Armstrong and Read 1998) also matters for growth, this dataset was downloaded from Alan Heston et al. (2006).

The actual donor pool depends on GDP data available. For some countries GDP is only available long after 1962. Therefore we face a trade-off between the size of the interval before the treatment and the number of donors available. Following Abadie et al. (2010), we maximize the length of the pretreatment period. Our donor pool contains almost 60 countries in most cases. For Namibia, however, we only have information about Namibia's GDP starting from 1970. Therefore, also the donor pool is slightly bigger. In Table 2.4 below we present the treatment states the time interval of the analysis and the size of the donor pool respectively.

Table 2.4: Treatment Years and Donor Pool Size

	Botswana	**Kenya**	**Lesotho**	**Malawi**
treat year	1987	1987	1991	1988
start year	1962	1962	1962	1962
end year	2007	2007	2007	2007
size donor	58	58	58	58

	Mozambique	**Namibia**	**Rwanda**	**S. Africa**
treat year	1990	1990	1983	1991
start year	1962	1970	1962	1962
end year	2007	2007	2007	2007
size donor	58	60	58	58

	Swaziland	**Uganda**	**Zambia**	**Zimbabwe**
treat year	1991	1982	1984	1985
start year	1962	1962	1962	1962
end year	2007	2007	2007	2007
size donor	58	58	58	58

Before we proceed to the main analysis we need to make some remarks about the countries considered in the analysis. Zimbabwe is currently one of the poorest countries in the world. In the pre-treatment period (prior to 1985) GDP per capita never exceeded 150$ (at PPP). Since the synthetic control group only allows interpolation from the convex hull and thus the elements of the vector $\mathbf{W}$ can only be positive, Zimbabwe's GDP per capita is too low to be matched by other countries. Therefore we are not able to find countries with similar growth paths and we are not able to analyze the effects of HIV. The

same holds true for Mozambique. Initial GDP in 1960 equals 153 $ (at PPP). Even though it rises to 289 $ by 1990 (the treatment year) no country can be found matching this growth path. Also Botswana is a critical candidate, however the reason is slightly different: its GDP went from 331$ in 1961 to 3130$ in 1986 (the year just before the intervention). This rapid growth was not matched by any other country in the world, and therefore we do not have any counterfactual for this country either. Two further countries are problematic and possibly have to be excluded: Rwanda and Uganda. Both of them have had wars with severe consequences for the population and for economic growth. The data reveal that Rwanda's GDP per Capita dropped by 60% in the year 1994 due to the effects of war and genocide, while Uganda's GDP dropped by 20% due to a war in 1979. Such a rapid decline in growth cannot be matched by other countries.[12] For all of these countries it is impossible to find suitable donors that match this behavior and thus the effect of HIV can not be identified, since it overlaps with different other circumstances.

Analyzing the remaining seven treated countries and their donor pools, we construct control countries. We build a synthetic control unit for each treated country using synthetic control groups. In the next Section we present the results of this optimization.

Since GDP per capita has a clear time trend in most countries, we focus on relative differences between treated and donor GDP in each year, instead of absolute deviations. Without this adjustment results are not comparable between countries.[13]

2.5 Results

In this Section, we present and discuss the main findings from our analysis. First, we devote our attention to the fit in the pre-treatment interval, and then we turn to the estimated effects.

[12]For the same reason we will exclude former members of the Soviet Union and other countries from the donor pool such as: Georgia (GDP per capita dropped by 55% in 1992), Lebanon (GDP drop of 52% in 1989), Armenia (drop of 50% in 1992), Serbia (drop of 37% in 1993), Latvia (drop of 35% in 1992), Moldova (drop of 35% in 1994) Tajikistan (drop of 33% in 1992), Bosnia and Herzegovina (drop of 30% in 1992), Azerbaijan (drop of 25% in 1992), Lithuania (drop of 21% in 1992), the Kyrgyz Republic (drop of 20% in 1994) and Bulgaria (drop of 17% in 1996) from the further analysis (as a possible placebo candidates).

[13]In Abadie et al. (2010) this poses only a minor threat since the variation of the outcome variable, cigarette packs within the US, is much lower than variation of GDP per capita in the World.

2.5.1 Pre-treatment interval: Match between treated units and synthetic controls

Tables 2.5 and 2.6 below display the match of the synthetic control group with the actual data for all seven countries we are analyzing. In Table 2.5 we display the results using variables from the theoretical literature, while in Table 2.6 we choose the variables from a vast array of variables and use the variables that minimize the root of the mean squared prediction error (RMSPE). The RMSPE is displayed in the last row and provides information on the fit of treatment and synthetic control in the pre-treatment period. Following equation (3.4) the prediction error represents the average difference in the logarithm GDP of treatment and synthetic control in each year. Since the GDP is measured in logs and not in levels the percentage deviation will be minimized. This criterion is preferable to the absolute deviation, since the GDP is usually growing over time and this way the growth path will be matched. Finally, in the third column of for each country we compare the estimates of the synthetic control unit with a simple average of the donor pool over the same time period.

Table 2.5: Means of Control and Treatment Group: Theory Based Predictors

	Kenya			**Lesotho**			**Malawi**			**Namibia**		
Predictors	Treated	Synthetic	Average	Treated	Synthetic	Average	Treated	Synthetic	Average	Treated	Synthetic	Average
GDP p Cap start	578.047	588.597	2619.675	219.832	279.479	3209.104	233.873	280.221	2619.675	2476.831	2394.095	3632.563
GDP p Cap mid	639.771	658.737	4200.542	384.195	368.761	4625.789	310.231	317.443	4200.542	2349.967	2522.410	4771.887
GDP p Cap end	838.129	820.440	6877.661	643.403	779.971	8864.206	419.903	520.463	7260.500	2803.402	3085.495	8232.726
Cap Form p Cap	137.648	151.599	1046.254	120.361	117.254	1302.181	64.267	59.426	1070.488	480.098	703.941	1302.838
Perc LForce	41.486	40.735	39.619	41.440	43.205	40.410	46.593	41.995	39.689	29.460	33.285	40.223
Human Capital	4.423	4.334	7.048	4.050	3.766	7.105	3.197	3.106	7.048	6.224	4.849	7.138
Pop Growth	3.610	2.833	1.788	2.185	2.186	1.678	3.151	2.224	1.777	2.986	2.862	1.675
Total Pop in M	13.348	12.357	50.939	1.257	324.200	52.877	5.335	172.100	51.433	1.024	19.263	56.831
Value Add Ser	45.308	47.300	43.392	43.107	28.475	46.137	37.420	28.246	43.772	44.598	44.172	46.787
Value Add Ind	19.103	18.801	30.795	21.007	23.774	31.340	17.847	17.605	30.802	44.247	36.202	31.866
Value Add Agr	35.589	34.000	17.774	35.886	47.751	16.512	44.733	54.149	17.644	11.155	15.175	15.816
Rural Pop	87.672	79.745	48.889	89.276	86.637	46.490	92.606	90.781	48.659	75.085	50.999	45.999
RMSPE	.	0.054	1.347	.	0.114	2.170	.	0.088	2.111	.	0.051	0.271
	South Africa			**Swaziland**			**Zambia**					
Predictors	Treated	Synthetic	Average	Treated	Synthetic	Average	Treated	Synthetic	Average			
GDP p Cap start	3295.960	3532.077	3209.104	414.260	566.892	3209.104	483.210	483.376	2619.675			
GDP p Cap mid	3940.744	4791.631	4625.789	1010.298	1129.152	4625.789	705.673	674.742	3920.849			
GDP p Cap end	5758.561	8501.042	8864.206	2523.079	2517.122	8864.206	807.678	902.464	5832.941			
Cap Form p Cap	1061.496	1479.724	1302.181	290.019	349.062	1302.181	171.975	140.408	984.688			
Perc LForce	27.531	30.698	40.410	32.048	37.898	40.410	36.852	37.886	39.436			
Human Capital	5.927	5.154	7.105	4.200	5.692	7.105	4.465	3.688	6.706			
Pop Growth	2.322	2.250	1.678	3.221	2.192	1.678	3.259	2.585	1.826			
Total Pop in M	26.933	17.962	52.877	0.588	92.553	52.877	4.546	6.078	49.471			
Value Add Ser	51.863	49.812	46.137	40.478	38.445	46.137	27.978	38.359	42.154			
Value Add Ind	41.382	41.547	31.340	32.718	32.625	31.340	52.734	26.667	30.653			
Value Add Agr	6.754	8.641	16.512	26.804	28.930	16.512	12.101	33.172	18.220			
Rural Pop	51.244	48.335	46.490	84.072	72.639	46.490	68.766	82.508	49.588			
RMSPE	.	0.084	0.352	.	0.118	1.147	.	0.057	1.294			

GDP p Cap stands for the GDP per capita at the beginning and the middle and the end of the pre-treatment period. Cap Form p Cap represents the capital formation per capita. Perc LForce measures the percentage of the inhabitants in the labor force. Human Capital is measured as average years of Schooling, Pop Growth represents the growth rate of the Total Pop in M(illions) in per Cent. Finally, our predictors include the value added by the service, industrial and agricultural sectors as percentages and the per Cent of inhabitants living in rural areas. RMSPE stands for the root of the mean squared prediction error. All predictors except for GDP are averaged over the pre-intervention period.

Table 2.6: Means of Control and Treatment Group: Empirical Predictors

Kenya				**Lesotho**				**Malawi**				**Namibia**			
Predictors	Treated	Synthetic	Average	Predictors	Treated	Synthetic	Average	Predictors	Treated	Synthetic	Average	Predictors	Treated	Synthetic	Average
GDP p Cap start	578.047	629.798	2748.996	GDP p Cap start	219.832	270.410	3275.372	GDP p Cap start	233.873	253.741	2698.547	GDP p Cap start	2476.831	3159.160	3869.680
GDP p Cap mid	639.771	608.290	4387.395	GDP p Cap mid	384.195	332.636	4693.540	GDP p Cap mid	310.231	305.212	4256.071	GDP p Cap mid	2349.967	3450.347	4989.836
GDP p Cap end	838.129	845.068	7069.874	GDP p Cap end	643.403	743.894	8886.832	GDP p Cap end	419.903	542.574	7086.283	GDP p Cap end	2803.402	6330.320	8521.207
Pop Growth	3.610	2.994	1.823	Area Tropics	0.011	0.304	0.257	Total Size Mkm2	0.111	1.099	1.372	GNP GDP ratio	86.699	91.640	98.068
Openness Measure	0.606	0.578	0.499	Latitude	-29.691	25.308	24.921	Openness Measure	0.614	0.165	0.422	Latitude	-22.104	6.720	23.819
Political Rights	5.330	4.973	3.130	Value Add Ind	21.007	27.185	31.666	Number Ethn Gps	6.000	7.438	4.037	RMSPE	.	0.042	0.312
RMSPE	.	0.044	1.381	RMSPE	.	0.106	2.193	Density 1960	26.659	15.166	99.129				
								Democracy Score	0.320	1.604	5.664				
								RMSPE	.	0.076	2.116				

South Africa				**Swaziland**				**Zambia**			
Predictors	Treated	Synthetic	Average	Predictors	Treated	Synthetic	Average	Predictors	Treated	Synthetic	Average
GDP p Cap start	3295.960	4297.585	4473.441	GDP p Cap start	414.260	592.247	3274.753	GDP p Cap start	483.210	531.547	2620.489
GDP p Cap mid	3940.744	5147.026	6406.770	GDP p Cap mid	1010.298	1148.863	4686.925	GDP p Cap mid	705.673	696.583	3910.450
GDP p Cap end	5758.561	8154.849	12482.447	GDP p Cap end	2523.079	2662.569	8825.189	GDP p Cap end	807.678	958.771	5702.235
Political Rights	7.000	6.356	5.895	KOF Glob Index	40.501	36.742	46.653	Perc Muslim	0.000	0.184	0.170
Latitude	-29.051	-21.844	30.070	LForce Female	40.405	35.450	35.194	Value Add Agr	12.101	24.283	18.723
Black Market	0.025	0.133	0.116	Political Rights	2.000	2.053	5.317	Latitude	-13.495	3.436	21.686
Tax Evasion	2.400	3.090	3.202	MASL	305.444	385.559	612.349	Political Viol	0.050	0.397	0.736
RMSPE	.	0.026	0.220	RMSPE	.	0.110	1.169	RMSPE	.	0.086	1.313

Area Tropics represents the percent of the land area in the tropics. Total Size Mkm2 stands for the total land size in million square kilometers. GNP GDP ratio is the ratio from GNP to GDP and is supposed to be a proxy for openness. The Openness Measure comes from the Ciccone and Jarocinski (2010) dataset. The KOF Glob Index represents a globalization index. MASL stands for meters above sea level. All predictors except for GDP are averaged over the pre-intervention period.

Looking at the theory-based predictors first, GDP per capita varies largely from country to country. While the average of the control group is around 3,000 PPP $ in 1961 the treated countries exhibit a GDP of around 500 PPP $ with a vast variation upwards (for example South Africa with 3,300 PPP $) as well as downwards (Lesotho's average GDP per capita was equal to 219 PPP$). GDP per capita roughly thirty years later (just before the treatment) is of course higher, for example Lesotho's GDP per capita increased to over 600 PPP $ and also the GDP per capita of South Africa reached almost 6000 PPP $. The average number of schooling years ranges from 3 (Malawi) to 6 (Namibia). Also sectoral composition varies largely from country to country. While most treated countries have around 30-40% of the workforce in agriculture, Namibia, South Africa and Zambia form exceptions with 11%, 7% and 12% respectively. Consequently these countries also have a lower share of rural population with 75% 51% and 68% respectively.

When we instead consider data-driven predictors, population growth is the only theory-based predictor that survives. Instead, political rights (Kenya, South Africa and Swaziland) and latitude (Zambia, South Africa and Namibia) gain importance. Moreover, openness seems to be an important predictor for Kenya and Malawi. Other important predictors are the sectoral composition (Lesotho and Zambia) and institutional quality measured by the democracy score (Malawi), tax evasion and the black market premium (South Africa).

Comparing the predictive power of the synthetic control unit with that of a simple average amongst donor countries, the advantage of the method becomes obvious. For GDP per capita we find very good synthetic control groups (less than 5% in terms of the root of the squared prediction error) for Kenya, Namibia and South Africa.[14]. All others have a prediction error higher than 5%, with Swaziland and Lesotho having the highest deviations around 10%.

[14]For comparison, it may be noted that Abadie et al. (2010) in their study find a RMSPE of about 3%, analysing sales of cigarette packs in the United States.

Table 2.7: Country Weights

	Panel A: Theory Based Predictors								**Panel B: Data-Driven Predictors**								Overall
Donor	Kenya	Lesotho	Malawi	Namibia	S. Africa	Swaziland	Zambia	Sum	Kenya	Lesotho	Malawi	Namibia	S. Africa	Swaziland	Zambia	Sum	Sum
Nepal	0	0.519	0.689	0	0	0	0.07	1.278	0	0.234	0.771	0		0	0	1.005	2.283
Indonesia	0	0	0	0	0	0.617	0	0.617	0	0.125	0.151	0	0	0.649	0	0.925	1.542
China	0.002	0.333	0.184	0	0	0	0	0.519	0	0.403	0.078	0	0	0	0	0.481	1
Sri Lanka	0	0.049	0	0	0	0.08	0.323	0.452	0	0	0	0		0	0.514	0.514	0.966
Madagascar	0.372	0	0.12	0	0	0	0.036	0.528	0	0	0	0		0	0.298	0.298	0.826
Mauritania	0.059	0	0	0	0	0	0.545	0.604	0.146	0	0	0		0	0	0.146	0.75
Algeria	0	0	0	0.166	0.536	0	0	0.702	0	0	0	0		0	0	0	0.702
Comoros	0.291	0.078	0	0	0	0	0	0.369	0.312							0.312	0.681
Malaysia	0	0	0	0	0	0.261	0	0.261	0	0	0	0	0	0.351	0	0.351	0.612
Nicaragua								0	0		0	0.596		0		0.596	0.596
Senegal	0	0	0	0.11	0	0	0.026	0.136	0.433	0	0	0		0	0.001	0.434	0.57
Philippines	0.176	0	0	0	0	0	0	0.176	0	0	0	0	0.272	0	0	0.272	0.448
Niger	0.101	0	0	0	0	0	0	0.101	0.109	0.238	0	0		0	0	0.347	0.448
Argentina	0	0	0	0	0	0	0	0	0	0	0	0	0.372	0	0	0.372	0.372
Netherlands	0	0	0	0	0.311	0	0	0.311	0	0	0	0	0	0	0	0	0.311
Iran	0	0	0	0.267	0	0	0	0.267	0	0	0	0		0	0	0	0.267
Chile	0	0	0	0	0	0	0	0	0	0	0	0.246	0	0	0	0.246	0.246
Bolivia	0	0	0	0.244	0	0	0	0.244	0	0	0	0		0	0	0	0.244
Peru	0	0	0	0.214	0	0	0	0.214	0	0	0	0	0	0	0	0	0.214
Australia	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0.2	0.2
Luxembourg	0	0	0	0	0	0	0	0	0	0		0.158	0	0		0.158	0.158
New Zealand	0	0	0	0	0	0	0	0	0	0	0	0	0.156	0	0	0.156	0.156
Morocco	0	0	0	0	0	0	0	0	0	0	0	0		0	0.152	0.152	0.152
Tunisia		0			0.101	0		0.101		0				0		0	0.101
Romania	0	0.021	0.007	0	0	0.042	0	0.07		0		0		0		0	0.07
Mexico	0	0	0	0	0.052	0	0	0.052	0	0	0	0	0	0	0	0	0.052
El Salvador	0	0	0	0	0	0	0	0	0	0	0	0		0	0.035	0.035	0.035
Sum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	7.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	7.00	14.00

Turning to the donor pool, Table 2.7 displays the actual weights given to the different countries when using the theory-based (Panel A) and data-driven predictors (Panel B), respectively, in order to replicate the trajectory of the treated unit.[15]. The differences in country weights between the two approaches are sometimes notable; for example in the case of South Africa, where none of the countries appearing in the theory-based set appears in the data-driven one. Even though countries like Argentina – having a high weight and allowing a much better fit (since the fit in general for South Africa is much better for the empirical donors) – for the data-driven predictors are available as donors, using theory based predictors it is not chosen by the estimator. Therefore the choice of predictors has a great influence on the choice of donor countries.

Nepal obtains the overall highest weight, while among the African countries Madagascar is the country with the highest weight.

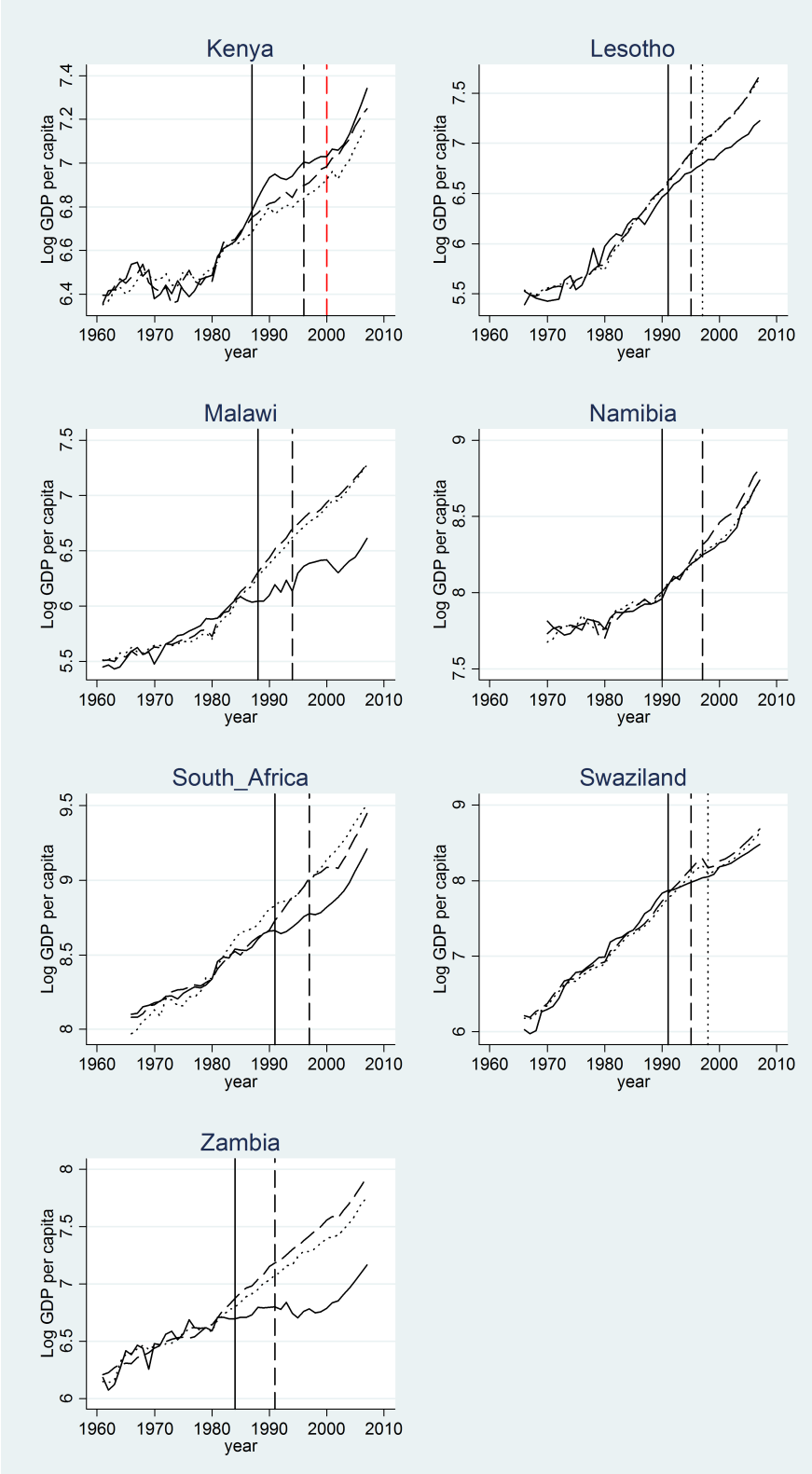
2.5.2 Post-treatment interval: Effects of HIV prevalence

We now turn to the effects of HIV prevalence on GDP. Our most striking result is actually that the estimated effects differ strongly from one country to another. Results can be found in Figure 2.3. The solid vertical line represents the cutoff, i.e., the year in which the HIV prevalence rate is thought to have crossed the 1% threshold. We also added some further reference lines: dashed lines indicate that HIV prevalence reached 10% and dotted lines signal that HIV prevalence transcends 20%. Red lines indicate that it crossed the respective prevalence rates in the opposite direction. In terms of GDP the solid lines represent the actual outocome of each country. Dashed and dotted lines represent the counterfactual, namely how the country would have behaved, based on our models, if there were no HIV. These counterfactuals are derived from weighting the donors as displayed in Table 2.7. The dashed line represents the weights derived from empirical (data-driven) predictors and the dotted line represents the counterfactual derived from theory based predictors. However, these results should be interpreted with caution since they only show the point estimates and do not tell anything about significance, which will be analyzed using placebo estimates in the next subsection.

For Kenya our figure shows that the actual GDP slightly exceeds both counterfactuals provided by the two different sets of predictors. For the other countries the result is the opposite: the actual outcome lies below the trajectories of the synthetic control group – thus suggesting a negative impact of the epidemic.

[15]Countries having a zero weight for all treatment units are dropped. Therefore the number of donors here do not correspond to the numbers of Table 2.4

Figure 2.3: GDP per Capita Trends, Treated Countries vs. Synthetic Controls



Note: Solid lines represent the actual development while the dotted and dashed lines represent the predicted outcomes by the synthetic control group for theory based and empirical predictors respectively.

Table 2.8 provides a clearer picture of the size of the effect. Following Abadie et al. (2010) we calculate the prediction error in the pre-treatment period (1st column) and for the period after the treatment (2nd column) and construct the ratio of the two RMSPE (last column). Moreover we estimate the average treatment effect in the third column, defined as the average difference between the actual GDP and its counterfactual (3rd column). Finally, in the forth column, we report the difference between synthetic control group and treatment country in the final period, i.e. 2007. Since the deviation is cumulative this variable also represents an important indicator of the total effect.

Table 2.8: Prediction Errors

Panel A: Theory Based Predictors

Country	RMSPE	RMSPEpost	Estimated Treatment Effect	TE final period	RMSPE ratio
Zambia	0.057	0.452	-0.414	-0.588	7.880
Malawi	0.088	0.488	-0.461	-0.665	5.550
South Africa	0.084	0.275	-0.268	-0.294	3.272
Kenya	0.054	0.135	0.133	0.171	2.522
Lesotho	0.114	0.264	-0.244	-0.419	2.325
Swaziland	0.118	0.088	-0.053	-0.151	0.739
Namibia	0.051	0.022	-0.011	0.002	0.432

Part B: Data-Driven Predictors

Country	RMSPE	RMSPEpost	Estimated Treatment Effect	TE final period	RMSPE ratio
South Africa	0.026	0.217	-0.211	-0.236	8.220
Malawi	0.076	0.529	-0.509	-0.669	6.940
Zambia	0.086	0.579	-0.538	-0.768	6.757
Lesotho	0.106	0.275	-0.255	-0.447	2.601
Namibia	0.042	0.088	-0.073	-0.083	2.116
Kenya	0.044	0.075	0.068	0.093	1.705
Swaziland	0.110	0.143	-0.125	-0.206	1.307

In both Panels, South Africa, Malawi and Zambia show the highest ratios, indicating that the difference between the counterfactual and the realized GDP after the treatment is considerably higher than before the treatment. Moreover, the estimated treatment effects are quite high for Malawi and Zambia, and measure between 40 and 50% of the GDP. Even though the treatment effect of South Africa around 20% is comparably small, the fact that the estimate occured with a very high precision for the data-driven predictors results in the highest RMSPE ratio of all, with an average deviation in the post treatment period around eight times as high as before the treatment. For the other countries these ratios are smaller and reach only around 2. In Swaziland and Zambia they are even smaller, indicating a larger deviation before the treatment than afterwards.

Contrasting the overall results from theory-based and empirical predictors we find that the RMSPE for the theory based models is higher (8% on average)

than for the empirical predictors (7% on average). This is also to be expected, since the latter approach has access to a larger set of predictors. Nevertheless, the treatment effects estimated by the two approaches are quite similar. Given this similarity, we focus on results based on empirical predictors for the remainder of this Section. The results of the robustness checks below are very similar for the theory based predictors. We will return to the theory based predictors in the next Section.

2.5.3 Placebo estimates

In order to assess whether the above differences are actually due to the treatment and not just random variation, we now provide an estimate of the significance of our findings using placebo estimates. We perform placebo estimations for the non-treated countries and compare the outcome these estimates to the countries severely affected by HIV. In order to get comparable estimates we randomly select the treatment year from the pool of treatment years in the severely affected countries. Figure 2.4 shows the results of this placebo analyses for all the years where we have a common support, i.e. data points more than 14 years away from the treatment year are not shown, in order to ensure comparability between the lines in the graph.

Figure 2.4: Placebo Estimates with respect to Treated Countries

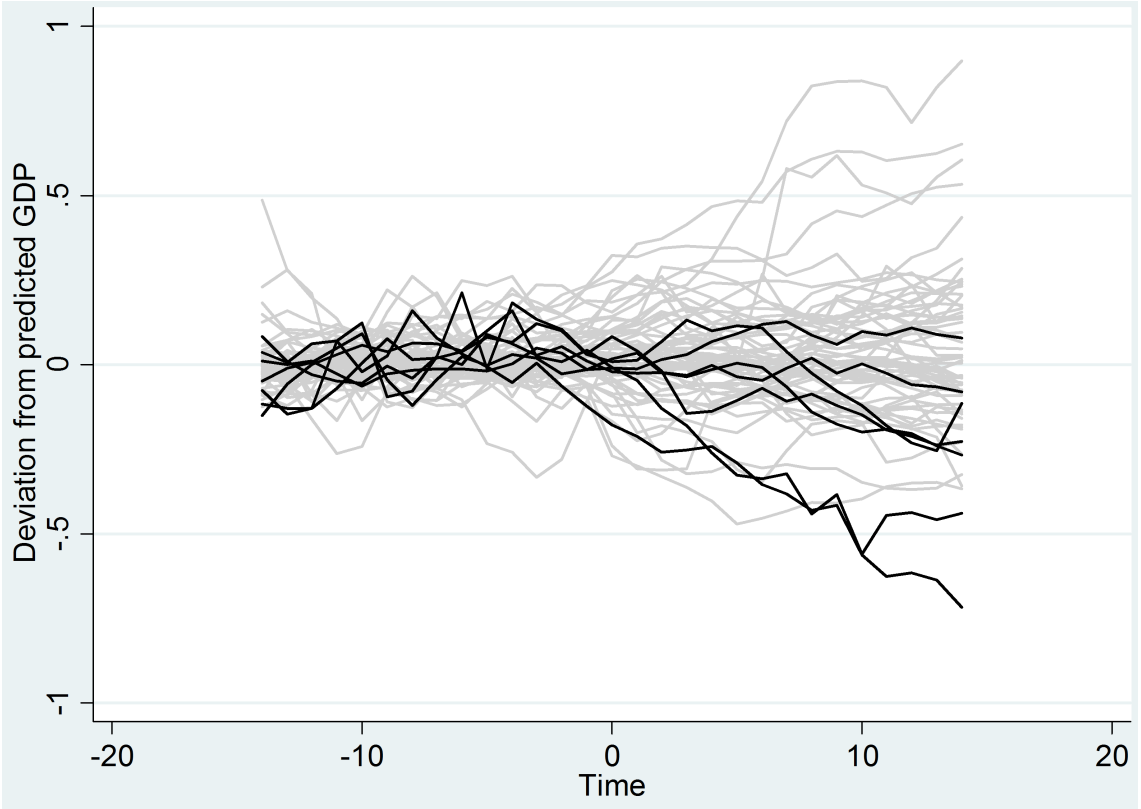


Figure 2.4 shows that Malawi and Zambia have suffered the largest estimated (negative) effect among all countries. Since there are 77 countries in the

graph this suggests that significance is reached at the usual 5% level. For the other countries the treatment effect is smaller. However, the effects might still be significant: For countries having a relatively bad fit in the pre-treatment period a large deviation post-treatment is only expected. Therefore the deviation prior to treatment needs to be considered as well. Following Abadie et al. (2010), we also provide the table of estimates with the RMSPE ratio.

Table 2.9: Prediction Errors for Empirical predictors: Placebo Analysis

	Country	Treatment	RMSPE	RMSPEpost	Estimated Treatment Effect	TE final period	RMSPE ratio
1	Mongolia	0	0.031	0.305	-0.295	-0.302	9.850
2	South Africa	1	0.026	0.217	-0.211	-0.236	8.220
3	Ireland	0	0.023	0.182	0.129	0.280	7.867
4	Canada	0	0.015	0.110	-0.102	-0.132	7.162
5	Malawi	1	0.076	0.529	-0.509	-0.669	6.940
6	Zambia	1	0.086	0.579	-0.538	-0.768	6.757
7	Comoros	0	0.057	0.310	-0.279	-0.535	5.463
8	Argentina	0	0.057	0.308	-0.286	-0.379	5.421
9	Malta	0	0.136	0.716	0.663	0.689	5.274
10	Spain	0	0.027	0.139	0.135	0.201	5.245
..							
32	Lesotho	1	0.106	0.275	-0.255	-0.447	2.601
..							
44	Namibia	1	0.042	0.088	-0.073	-0.083	2.116
..							
52	Kenya	1	0.044	0.075	0.068	0.093	1.705
..							
63	Swaziland	1	0.110	0.143	-0.125	-0.206	1.307
..							
77	Pakistan	0	0.097	0.049	0.001	-0.031	0.503

According to the Table 2.9, in which countries are ordered by their RMSPE ratio, South Africa, Malawi and Zambia rank second, fifth and sixth respectively indicating a significance of 2.5, 6.5 and 7.8 per cent respectively. The next treated country, Lesotho, has rank 34 and is thus far away from being significant according to this definition.

Summing up, there is substantial heterogeneity in estimated effects. For one group of countries, we find no effects of HIV on GDP per capita (Botswana, Lesotho, Kenya and Namibia) – while for South Africa, Malawi and Zambia we find effects that are quite large and also significant.

In order to check the robustness of our findings, we performed additional robustness checks not included in the current version of the paper. The synthetic control group estimator uses the mean in order to replicate a country, while the variance is not taken into account. For example a middle income country could be replicated by another country of middle income or by the mean of a high income and a low income country. The results in the post-treatment interval might depend upon whether rather similar or rather different countries are chosen by the estimator. In order to prevent that, we performed a robustness check including only countries in the donor pool that are close enough, both in terms of GDP per capita and population size. Using this restriction affected our results only slightly. While South Africa and Zambia ended up having a

higher RMSPE ratio, the ratio of Malawi decreased from 6.9 to 4.5. Another issue is that our estimator might be overfitting. Given the fact that we always include the first, the middle and the last year of the pre-intervention period of the dependent variable among the predictors, might be too restrictive. In accordance with theoretical models we only include the variable at the beginning (t=0). This increases the RMSPE ratio in Zambia and Malawi and slightly decreases the ratio in South Africa from 8.2 to 6.7. However, based on our previous ranking (Table 2.9) all treated countries would still be among the top six. Summing up, we see quite a robust treatment effect for Zambia, Malawi and South Africa, while for other countries we find many placebos with effects of similar magnitude.

2.6 HIV prevalence and growth

In the previous Section, we have identified negative effects of HIV in some countries, while for other countries also severely affected by the pandemic, there seem to be no effects. We now analyze how these estimated effects relate to the HIV prevalence rate and also to some key economic and public health indicators which can be suspected to be important mediators of the relationship. Thus, we apply the synthetic control group estimator (with theory-based and empirical predictors) to all countries in our sample (i.e. mildly and strongly affected countries). Then we employ regression analysis on the generated data to estimate the following equation:

$$\hat{\alpha}_{it} = c + \beta Z_{it} + \gamma R_{it} + \varepsilon_{it} \tag{2.5}$$

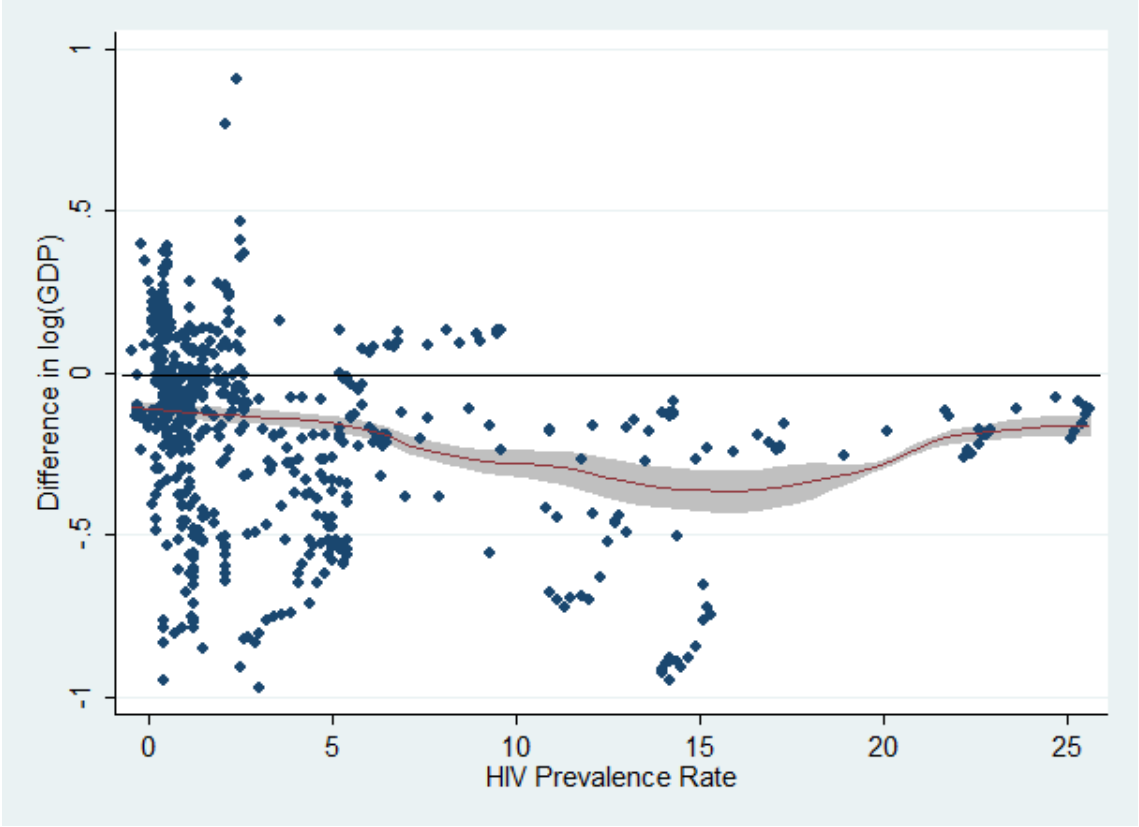
where, $\hat{\alpha}_{it}$ is the estimated treatment effect (the difference between actual and predicted GDP per capita), R_{it} represents HIV prevalence in country i a given year, Z_{it} are some other variables that might influence the difference such as political violence. We estimate the equation concentrating on the time interval after the intervention. Moreover, we weight observations by the performance of the estimator in the pre-treatment period (using the square of the deviation between the treated country and its synthetic control). In total, we have 551 (32 severely affected countries) and 658 (40 severely affected countries) country-year observations for the theory-based and the empirical predictors respectively.

First we present the relationship between HIV prevalence and relative GDP loss in each country (for the empirical predictors). Figure 2.5 below shows all single observations, each datapoint representing one country.

Figure 2.5 shows that there appears to be an approximately linear, and statistically significant, negative relationship between the level of HIV prevalence and the estimated effect in terms of GDP. This will be confirmed in the regression analysis below (Table 2.10).

In the first model (1) of each Panel we only include the HIV prevalence rate and year dummies as control variables. In the next step we vary the time lag of HIV. We find that a time lag of seven years leads to the highest explanatory power, measured by the Wald chi square indicator. This result is consistent with the fact that the average incubation time is around seven years as well. No higher orders than the linear term turned out significant. In a third step we include a dummy variable for African countries (column 3). Although the Africa dummy is significant in most specifications our estimate of the treatment effect is hardly altered. So in our estimates we find an Africa effect, but it appears not to bias the influence of HIV prevalence. In terms of

Figure 2.5: Local Polynomial of GDP on lagged HIV



the size of the effect of the HIV prevalence rate, we find that the GDP decreases between 1.1 and 2% for each additional percentage point of prevalence.

In the columns 4 and 5 of each Panel we control for other variables in order to identify important mediators. Here we concentrate on key economic and public health indicators and other variables that were chosen as important predictors for the data-driven synthetic control group estimator (see Table 2.6).[16] We find that land area in the tropics are an important mediator leading to larger treatment effects. The same holds true for political rights. For land size and pre-intervention population growth on the other hand we find no effect, even though our estimates are very precise.

[16] Although the latitude has been chosen as an important predictor including it as a linear term in a regression context, would have no meaning economically.

Table 2.10: Correlations between GDP and HIV

	Panel A: Theory Based Predictors					Panel B: Data-Driven Predictors				
	(1) Diff. in GDP	(2) Diff. in GDP	(3) Diff. in GDP	(4) Diff. in GDP	(5) Diff. in GDP	(1) Diff. in GDP	(2) Diff. in GDP	(3) Diff. in GDP	(4) Diff. in GDP	(5) Diff. in GDP
L.HIV Prevalence Rate	-0.0123***					-0.00601*				
	(0.00215)					(0.00311)				
L7.HIV Prevalence Rate		-0.0183***	-0.0171***	-0.0197***	-0.0201***		-0.0105***	-0.0110***	-0.0132***	-0.0197***
		(0.00256)	(0.00245)	(0.00227)	(0.00215)		(0.00373)	(0.00375)	(0.00319)	(0.00269)
Africa Dummy			-0.232***	-0.304***	-0.396***			-0.0426	-0.203***	-0.370***
			(0.0359)	(0.0343)	(0.0605)			(0.0326)	(0.0395)	(0.0575)
Land Area in Tropics (Per Cent)				-0.247***	-0.263***				-0.242***	-0.331***
				(0.0467)	(0.0412)				(0.0719)	(0.0815)
Land Size (Area in M km2)				-0.00945	-0.0135				-0.00333	0.00875
				(0.0498)	(0.0260)				(0.0605)	(0.0395)
Political Rights					-0.0338***					-0.0442***
					(0.00979)					(0.0112)
Population Growth (Per Cent)					0.0575**					0.0454
					(0.0291)					(0.0323)
Constant	-0.00436	0.0149	0.193***	0.392***	0.480***	-0.0622***	-0.0488**	-0.0262	0.202**	0.481***
	(0.0175)	(0.0181)	(0.0358)	(0.0471)	(0.0849)	(0.0208)	(0.0212)	(0.0264)	(0.0798)	(0.113)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	551	551	551	551	551	658	658	658	658	658
Wald Chi^2	74.98	94.81	142.60	295.03	460.73	51.20	59.16	59.73	90.19	284.36

The table represents the regression results of a random effects model with the yearly deviation from the counterfactual as a dependent variable. 32 (theory based predictors) and 40 (data-driven predictors) severely affected countries and all years after treatment year up to 2007 are included. Observations weighted by the inverse of the root mean squared prediction error (RMSPE) in the pre-intervention period. Standard errors corrected for unit root in parentheses. Political rights come from La Porta et al. (1999) and population growth Mundial (2011) from the year 1980; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.7 Discussion

In this Chapter, we estimate the effects of HIV prevalence on GDP in the twelve Sub-Saharan countries most heavily affected by the pandemic. We employ the synthetic control group approach (Abadie and Gardeazabal 2003), which provides a novel data-driven econometric method for case studies. With the help of this approach, we are able to find fairly good control groups for most countries in our sample. However, for Mozambique and Zimbabwe the method also shows its limits. Due to the fact that their GDP per capita are lower than in all other countries, we are not able to construct a good control group for them. For Botswana, on the other hand, the growth path is to steep to be replicated by any other country. Moreover, the method does not allow for disentangling the effects of different events that coincide with the intervention. For Rwanda and Uganda, political change occurs in the same time interval as the pandemic, which reduces the credibility of our estimates for these countries. Despite these problems, we are able to find fairly good control units for most treated countries in the interval before the intervention.

Our results suggest that HIV left substantial effects in South Africa, Malawi and Zambia, while for Lesotho, Swaziland, Kenya and Namibia we are not able to identify significant effects on GDP per capita. Concentrating on the countries where we find effects at the per capita level, our results suggest that those countries lost 24% to 77% in terms of GDP per capita as compared to the counterfactual without HIV. For Zambia we find the largest effect of the pandemic, with an cumulative impact of about 77% of GDP from 1984 to 2007. Considering that Zambia has a 24-year history of HIV, this means that on average Zambia had a growth rate which is 5.5% lower per year than what would have been achieved without HIV. The second largest effect, can be found in Malawi, where the effect measures around 67% of GDP from 1988 to 2007 (or a 4.5% reduction in the growth rate). Finally, for South Africa we find the smallest effect with around 23% of GDP from 1991 to 2007 or a 1.5% reduction of the growth rate.

In order to analyze the effect heterogeneity more closely we perform a regression analysis, where the difference in economic growth between treatment and control group was the dependent variable. This allowed us to take a step into the black box, seeing what contributed to the economic consequences of HIV prevalence.

We find that the effects of HIV on GDP arise with a time lag of around seven years. This time interval matches the average incubation time without medication. Since initially the disease is "'sleeping"', the infected individuals don't exhibit symptoms. Therefore it seems reasonable that economic consequences can be measured only after the outbreak of the disease. The effect we

find is is highly significant and thus suggests that on average there are large effects of GDP on HIV. Quantitatively we find that each percentage point increase in prevalence leads to a decrease in GDP of around 2%.

Moreover, the regression analysis shows that African countries exhibited more severe economic consequences. This result is expected as well. However, even when including an Africa dummy the point estimate of the prevalence rate remains unaffected.

Furthermore, we were able to identify important mediators. The first mediator is a higher percentage of land area in tropics. According to Gallup et al. (2001) countries in tropics develop rather poorly compared to other countries, since diseases spread more rapidly. People infected with HIV, usually do not die from AIDS itself, but their immunity system is weakened by HIV making them more vulnerable for other diseases. Therefore a region prone to diseases might be especially dangerous for individuals, whose immunity system has already been weakened by HIV.

For population growth we find mixed effects. The claim of Young (2005), that HIV had also positive consequences since it stopped excessive population growth in African countries, might lead to the conclusion that a high population growth will lead to less severe consequences of HIV. Our data shows that population growth works as a mediator and thus reduces economic growth. However, in order to verify the claim of Young (2005) one would still need to confirm that HIV affects fertility and population growth which will be analyzed in the next Chapter.

Finally, we also find that more political rights exacerbate the economic consequences of HIV. This finding might seem a bit surprising at first glance, however the logic is just the same as with population growth, although in the opposite direction. Countries with more political rights are found to have a higher growth potential (La Porta et al. 1999). Therefore there is potentially more economic prosperity to be disrupted by HIV prevalence in these countries.

In the next Chapter we analyze the effects on demographic variables in order to gain more insight in the different channels.

CHAPTER

3

Demographic Consequences of HIV **

3.1 Motivation

As we saw in the previous Chapter the treatment effects of HIV appear quite heterogeneous. This is in line with previous literature. Many papers estimate the impacts on economic growth and results vary widely. While some studies find quite large negative effects (see for instance Gaffeo 2003, who estimates a reduction in the yearly growth rate of 1.2-1.7 percentage points), some studies like Bloom and Mahal (1997b) find no effects, and Young (2005) even predicts positive effects. This ambiguity might result from effects working through many different channels and different channels dominating in different countries.

In this Chapter we analyze the impact of the HIV/AIDS pandemic on demographic variables for the Sub-Saharan region. Estimating the demographic impact of HIV restricts the possible channels and therefore helps to clarify the heterogeneity in terms of economic consequences. For instance Young (2005) suggests that HIV leads to a reduction in fertility and therefore fosters economic growth for future generations. In order to assess the validity of this claim it is important to obtain a clear estimate of the effects of HIV on fertility.[17]

Since HIV/AIDS is a lethal disease a substantial disruption of demographic

**This Chapter is based on Karlsson and Pichler (2012a).

[17]Several later studies – such as Kalemli-Ozcan (2012) – reject the result of Young (2005), however in this paper we combine an analysis with other demographic variables in order to obtain a clearer relationship.

variables such as life expectancy and the mortality rate can be expected. However, a clean estimate of the causal impact is complicated by the fact that important secular trends in several demographic variables were observed before the pandemic hit. Furthermore, life expectancy in African countries is quite low in general, compared to other countries with similar development. Therefore it is important to assess how much of the difference is due to HIV and to which degree other conditions are responsible.

Moreover, HIV will very likely change incentives for engagement in sexual activity and thus there might also be an effect on fertility. However, different mechanisms imply different results. On the one hand women who know about their infection might not want to infect their children and partners and thus fertility could be reduced. Moreover, there exists evidence that HIV biologically affects fertility through increased rates of miscarriage and stillbirth (see for instance Fabiani et al. 2006). On the other hand HIV might also lead to positive effects in the birth rate. Chen (2010) suggests that higher life expectancy in general leads to a decreased fertility since parents might prefer to concentrate their investments on fewer children. HIV might reverse this pattern since the decrease in life expectancy might lead to an increase in fertility. Estimating the effect of HIV on fertility helps to identify which mechanism prevails.

This is the first analysis in the literature to focus on the three outcome variables life expectancy, death and birth rates. Other papers with a demographic focus concentrate on direct effects and thus estimate the effects on the mortality and life expectancy (see for instance the publication of the UN Population Division 2004). A second often considered relationship is the behavioral effects on fertility (see Kalemli-Ozcan 2012, for a recent study on this topic). Analyzing both direct and indirect demographic effects together is useful because the effects are interrelated. The mechanism considered above with lower life expectancy leading to to higher fertility can only take place if there is in fact a large enough drop in life expectancy.

As in the previous Chapter, we focus on the effects of HIV/AIDS in the twelve most severely affected countries. Also the method will be similar. However, in this Chapter we extend the approach of Abadie and Gardeazabal (2003) to three variables.

Our results show that HIV decreased life expectancy by 15 years on average, while the death rate was increased by seven additional deaths per 1,000 inhabitants. In terms of the birth rate however, we do not find any effect.

This Chapter is structured as follows: First an overview of impact of HIV on demographic factors will be given. In Section 3.3 the synthetic control method for comparative case studies of Abadie and Gardeazabal (2003) will be

presented and discussed, followed by the description of the data and variables used in Section 3.4. In Section 3.5 we will describe how we adopt the method to our setting with three outcome variables. The results are presented in Section 3.6. Finally, an evaluation of the achieved results will be given in Section 3.7.

3.2 Literature review

We already saw how the disease affected different parts of the world in Section 2.2 in the previous Chapter. In this Section we will briefly review the literature on the effects of HIV on demographic variables.

The increase in life expectancy was in the past driven by a decline in mortality rates which came along with improved medical care and consequently improved health (Cutler et al. 2006). This increase in life expectancy was therefore strongly associated with economic growth even though the relationship between the two variables is a complex and dynamic one (Fogel 2004). Chen (2010) shows in a theoretical model that a higher life expectancy will in general lead to a decline in the fertility rate and a higher educational level due to a quality-quantity trade-off concerning the number of children and their education. But he also considers a possible steady-state situation with exogenous mortality rates, where the demography of a country is caught in a "poverty trap" with characteristically high fertility rates, low educational level and low life expectancy. Becker et al. (1990) also consider two steady-state situations, where one equilibrium is characterized by a low fertility rate, increasing investments in education and physical capital and the second steady-state is found with a high fertility rate and consequently low investments in human and physical capital. This is referred as the "Malthusian" equilibrium which can only be changed with a strong investment policy. The fertility rate, mortality rate and educational level are mostly seen as decisive factors for life expectancy (see e.g. Becker et al. 1990). Even if it is not really known how these determinants contribute to changes in life expectancy it is likely that an external shock to the determinant "mortality" would probably cause a reaction in all other variables. The HIV/AIDS pandemic can be seen as such an external shock.

Several studies contemplate the relation of demographic determinants such as fertility rate to the HIV prevalence rate. Fortson (2009) reports only a small effect of HIV/AIDS prevalence rates on fertility rates when comparing 12 different Sub-Saharan countries. She argues however, that if educational levels are taken into account, the relation between HIV/AIDS prevalence rates and fertility becomes evident. Fink and Linnemayr (2008) also consider how education affects the effect of HIV on fertility. They find a weak and statistically

insignificant positive correlation between the HIV/AIDS prevalence rate and the fertility rate for the overall data, but if one considers the data in groups with different educational levels a positive effect of HIV/AIDS prevalence on fertility of non-educated mothers and mothers with primary schooling is found. But there is a negative effect of HIV/AIDS prevalence on fertility of mothers with secondary and higher school education. The social transformation therefore leads to smaller prevalence amongst the better educated. Thus education has an effect on the average HIV/AIDS prevalence rate in one or another way. Whether the HIV/AIDS prevalence rate influences the decision for human capital investments, or the fertility rate in HIV/AIDS infected areas is influenced by the educational level of the parents, cannot be solved here. Durevall and Lindskog (2011) also investigate the effect on fertility and find mixed results in Malawi. They find a relation between the age of mothers and their desired number of children. Their study shows that younger women increase their desired number of children while older women decrease it. The mechanism is not directly traceable, but it is believed to have its cause in the prevention of giving birth to HIV-infected children. Moreover, Young (2005, 2007) identifies a large negative effect of the epidemic on fertility, whereas other studies (Kalemli-Ozcan 2012, Juhn et al. 2008) report a much smaller or even positive effect. Summing up, there is remarkable disagreement as to the effects of HIV on fertility (Durevall and Lindskog 2011).

A further related subject is how investment in human capital is affected by HIV/AIDS prevalence rates. Since there is a strong correlation between fertility rates and educational attainment, several studies focus on the role of education with respect to HIV/AIDS prevalence rates. Fortson (2011) finds that the HIV/AIDS prevalence rate is directly negatively correlated to years of school attendance. Her analysis shows that a prevalence rate of 10 percent lowers the average years of school attendance by 0.5 years. She resumes that the lower investment in human capital is a consequence of the higher mortality risk and the associated drop in investment.

A further parameter which is obviously affected by the HIV pandemic is the mortality rate since HIV/AIDS is a deadly disease. In 2004 about 3.9 percent of all deaths worldwide were caused by AIDS. While in most parts of the world this fraction was around 1 percent, it was 15 percent in Sub-Saharan Africa. An even more drastic picture emerges if one takes a look at age groups. Among adults aged 15-59 the dominant cause for death in Sub-Saharan Africa in 2003 was AIDS (e.g. 85 percent in Botswana and 61 percent in Uganda). The projected percentage of deaths caused in 58 countries most affected by HIV will remain at a level of around 16-18 percent of all death for adults aged 15-59. The group of adults between 15-59 years additionally makes up for 86

percent of all AIDS victims in 2007 (Bongaarts et al. 2010). This makes it especially dramatic for the economic development of the Sub-Saharan region because these deaths reduce the labor force of this area considerably. The death rate connected with HIV is therefore not a direct relationship, but is also relevant for the demographic composition.

In summary, it is not surprising that HIV/AIDS has had an effect on life expectancy and death rates in this area (UNAIDS 2010). However, the size of this impact and the effect on other key demographic indicators such as the birth rate is still an open issue. The estimation of magnitudes of effects is difficult considering secular trends and the selection effects inherent in HIV mortality. Thus, there is no obvious counterfactual available, against which the impact of the pandemic may be assessed.

We use the synthetic control group method introduced by Abadie and Gardeazabal (2003) and already used in the previous Chapter to generate such a counterfactual. The main challenge in our setting is that the method so far was only used with respect to one single outcome variable. We have identified three variables of interest, namely life expectancy, death and birth rate. Therefore we extend the approach introduced by Abadie and Gardeazabal (2003) to several dependent variables.

As in the previous Chapter we will use the HIV data from Oster (2007). The other variables used here were taken from different sources and come with the same problems as the HIV/AIDS prevalence rates. Data for birth rates and death rates were taken deployed from the World Bank datasets (Mundial 2011) and data on life expectancy was collected from different sources and compiled by Rosling (2012).

Finally, we use the same thresholds already defined in the previous Chapter (see Section 2.2.1). Thus, we will refer to a country as *strongly affected* if the estimated prevalence rate is above 1 per cent. Likewise, we refer to countries with prevalence rates above 10 per cent as *severely affected*. The rest of the countries, i.e., those that have never crossed the 1 per cent threshold, will be referred to as *mildly affected*.

3.3 Econometric Approach: Synthetic Control Groups

The synthetic control group method tries to find matches from a donor pool in order to minimize the distance for *one* outcome variable. This poses a problem because the minimization occurs along one dimension, while we have three variables of interest, namely life expectancy, death and birth rates. In order

to circumvent this problem we extend the method so that it can be applied to multiple variables. Before we come to a more detailed description of the extension method and the weighting in Section 3.5, we revise the standard assumptions for the method in this Section and present the data in the next Section. In what follows the dependent variable Y will be a vector consisting of the variables life expectancy, birth and death rate.

Social scientists and economists often try to measure the consequence of a specific action or event (e.g. law, tax, disaster) on a certain outcome. In order to estimate the effect that is attributable to a specific intervention, comparative case studies are often used. This technique employs control groups which are unaffected by the analyzed intervention and compares them with the affected group. The control groups themselves are supposed to be comparable to the affected group before the intervention started and the differences between the control groups and the treated one are used to calculate the total change caused by the intervention. Although this method provides a good approach for the calculation of the intervention's effect, it comes with two methodological problems. First, the selection of the control groups, which is normally based on some kind of affinity to the affected group. Second, the uncertainty about whether the true behavior of the treated group in absence of the treatment would resemble the outcome of the counterfactual (Abadie et al. 2010).

In order to address these problems the synthetic control group approach (for comparative case studies) offers a useful tool. The selection of the control groups is based on data-driven procedures which reduce the bias caused by subjective criteria of the researcher since the researcher has limited influence on the selection of the control groups. Additionally, the synthetic control group approach provides a safeguard against extrapolation – which is one main weakness of regression-based methods. This is achieved by the construction of the synthetic control group through the calculated weighted average of the chosen donor groups, where weights are restricted to be non-negative. Thus, the synthetic control group is the weighted average of comparable control groups which increases the reliability of the prediction in comparison to traditional regression methods (Abadie et al. 2010). An additional strength of this method is that no information on post-intervention outcomes is needed to design the study. Thus, there is a much lower risk that the study design is biased.

3.3.1 Assumptions

We now present the main assumptions needed for the analysis. Since the HIV epidemic reached critical levels in different countries at different points in time, each affected country needs to be analyzed separately. Thus, in our notation

below we proceed as if there were only one single treated unit and J further control units.

We denote by Y_{it}^{NI} the outcome that would have been observed in country i at time t in absence of the HIV epidemic. Also, we denote by T_0 the last pre-intervention period, i.e.

$$T_0 = \inf \{t | R_{1t} \geq 0.01\} - 1$$

where R_{1t} denotes the HIV prevalence rate in the affected country.

Moreover, let Y_{it}^{I} be the outcome for the affected country i at time t, where the epidemic had started to take off at time $T_0 + 1$. Since HIV is unlikely to have had an effect on the outcome variables before the outbreak of the pandemic, we also have $Y_{it}^{I} = Y_{it}^{NI} \forall t = 1, \ldots, T_0, \forall i = 1, \ldots, J + 1$.

Next, define $A_{it} \equiv Y_{it}^{I} - Y_{it}^{NI}$ as the effect of the HIV epidemic for unit i at time t, and let D_{it} be an indicator which takes on the value one whenever the HIV epidemic has crossed the 1 per cent threshold: $D_{1t} = \mathbf{1}(t > T_0)$. Thus, the observed outcome for unit i at time t equals

$$Y_{it} = Y_{it}^{NI} + A_{it} D_{it} \tag{3.1}$$

Abadie et al. (2010) suggest that the untreated value Y_{it}^{NI} can be described by a factor model given by the following equation:

$$Y_{it}^{NI} = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \tag{3.2}$$

where δ_t is a time effect common to all units, θ_t is a vector of possibly time-dependent coefficients, λ_t is a vector of unobserved common factors, and μ_i is a vector of unknown factor loadings.

A further assumption which is needed is that the outcomes of non-treated countries are unaffected by the intervention. There are several ways in which this assumption could be violated. Firstly, there is obviously the risk of contagion, to the extent that the countries of the control group also become affected by the pandemic. For example, Oster (2007) delivers strong evidence that trade between countries elevates incidence rates. However, this possibility has been eliminated by considering only countries which have had very low prevalence rates throughout.

It is furthermore desirable that there is no other extraordinary impact on the variables forming the vector after time period T_0 neither for the regarded country nor for the donor countries. These impacts could be hunger crises, natural disasters, wars etc. It could of course be argued that, since the comparison units – the synthetic control groups – have been defined so as to closely

mimic the behavior of all relevant variables in the treated country, including severe and unexpected exogenous shocks. However, inference from the method is based on small samples and for this reason, the occurrence of an exogenous large-scale shock can lead to erroneous conclusions. Hence, major disasters appearing after the treatment started need to be considered, and countries exposed to such shocks should be excluded from the analysis as a sensitivity check. Information on these major shocks can normally be obtained from public sources like UN reports or media archives.

A possible bias in this study can also come from migration to or from the included countries. This happens if the life expectancy (or birth rates) of the immigrants in comparison to the life expectancy of the immigration country differs strongly and the number of immigrants is high and thereby significantly affecting the population growth in these countries. However, the outflows from Sub-Saharan Africa are likely to be too small to have had an impact on the receiving countries' demographic indicators. For example, the total stock of immigrants from Sub-Saharan countries in OECD countries was around 3.9 million in 2002, working out at less than 0.5 per cent of the total population (OECD 2005).

3.3.2 Implementation

The synthetic control method involves estimating two matrices: $\mathbf{V}$ is the weighting matrix determining the relative predictive power of various outcome variables Z_i and of the outcome variable itself. The vector $\mathbf{W}$ is a vector of non-negative weights given to the J control countries.

The criterion minimized is given by

$$\left\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0\mathbf{W}\right\|_{\mathbf{V}} = \sqrt{\left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0\mathbf{W}\right)' \mathbf{V} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0\mathbf{W}\right)} \tag{3.3}$$

where $\bar{\mathbf{X}}_j$ is a vector of averages over the pretreatment period of elements of Z_i and Y_i, for treated and control units, respectively. This will give us an optimal country weight matrix among all diagonal positive definite matrices depending on the variable weight ($W^*(V)$).

The predictors V are chosen such that the average distance from of the outcome variable, i.e. the root of the mean squared prediction error is minimized in pre-intervention periods:

$$RMSPE = \sqrt{\frac{\sum_t \left(Y_{1,t} - \mathbf{Y_{0,t}}\mathbf{W}^*(\mathbf{V})\right)^2}{N}} \tag{3.4}$$

where N represents three times the number of years in the pretreatment period (since the outcome variable consists of three variables).

3.4 Data and Variables

For our analysis we use data from various sources. As already mentioned above the main source of our HIV data is Oster (2007). All other variables were obtained from Mundial (2011). Furthermore we used data from Rosling (2012) to enrich this dataset and replace missing values.

Our dependent variable is a vector based on crude birth and death rates (per 1,000 inhabitants) and average life expectancy (at birth). We use crude birth and death rates instead of adjusted mortality and total fertility rates, because these measures provide the largest dataset in terms of countries we are able to consider. Besides, it is more straightforward to compare crude birth and death rates. In addition, we control for GDP per capita and human capital measured as years of schooling.[18] We also considered controlling for the population structure; however, data availability was the main hurdle.

As Section 3.3.3 reveals the choice of variables included in $\mathbf{X}$ will influence the countries the estimator uses to replicate the treated country and will thus also determine the outcome of the estimation. Therefore it is important to include demographically meaningful predictors such as life expectancy, death and birth rate. While the death rate and life expectancy are directly connected to the HIV/AIDS pandemic, the birth rate is only indirectly connected to it through various behavioral responses. Moreover, we saw in Section 2 that the living standard which we measure by the GDP per capita is likely to influence population growth. The last independent variable used in all calculations is the average years of school attendance. This variable is influencing not only economic development and birth rates, but is also a vital factor for strategies for HIV prevention and knowledge about the disease as described above.

Combining all these variables from various datasets, we are able to construct a balanced panel dataset for 103 countries (12 treated and 91 in the donor pool, 11 of them are African countries) from 1960 to 2008. The median year of the intervention (when the HIV prevalence rate increases to more than 1%) is 1987, with minimum of 1982 and maximum of 1991, thus giving us around 27 pre-intervention years and 22 post-intervention years.

The synthetic control group estimator builds the average of several control countries in order to replicate a treated country. However, the variance is not taken into account. For example a country with a life expectancy of 50 years could be replicated by another country with a life expectancy of 50 years or of two countries with a life expectancy of 30 and 70 years respectively. The results might depend upon whether rather similar or rather different countries are chosen by the estimator. Abadie et al. (2010) recommends to use rather similar

[18] This last variable comes from Lutz et al. (2007) and Barro (2001). The full list of variables and data sources may be found in Appendix B.

donors. However, we find that the fit before treatment depends too strongly on the donor, therefore we extend the estimator by considering different sets of donor pools. First we only consider donors that are very close in terms of all three outcome variables (1.5 standard deviations), then we stepwise include more and more donors up to the full sample of donors. Finally, the donor pool with the lowest RMSPE (defined in equation 2.4) was chosen. Table 2.4 below shows the number of donors considered for each group. South Africa and Namibia lie somewhat in the center of the distribution and thus have a high number of close potential donors minimizing the RMSPE. On the other hand Malawi, Mozambique and Rwanda have quite low life expectancies and high death and thus lie in the tails of the distribution and therefore more donors do not improve their RMSPE.

Table 3.1: Treatment Years and Donor Pool Size

	Botswana	**Kenya**	**Lesotho**	**Malawi**
treat year	1987	1987	1991	1988
start year	1960	1960	1960	1960
end year	2008	2008	2008	2008
size donor	48	50	47	27

	Mozambique	**Namibia**	**Rwanda**	**S. Africa**
treat year	1990	1990	1983	1991
start year	1960	1960	1960	1960
end year	2008	2008	2008	2008
size donor	27	46	11	46

	Swaziland	**Uganda**	**Zambia**	**Zimbabwe**
treat year	1991	1982	1984	1985
start year	1960	1960	1960	1960
end year	2008	2008	2008	2008
size donor	49	32	23	25

The table shows the considered years for each countries. Our starting year is 1960, and our data ends in 2008. Treatment years, i.e. when HIV prevalence passes the one percent level, differ between countries. In the last row of each country we display the size of the considered donor pool for each country. As suggested by Abadie et al. (2010) we only include donors that are quite comparable in terms of the three outcome variables. The whole donor pool consists of 91 countries. Out of this pool we first choose countries that are close in terms of the outcome variables. Afterwards, we stepwise include more possible donors and then use the donor pool that minimizes the RMSPE – defined in equation (2.4) – for our estimation. In this way we reach the displayed donor pool for each country.

Analyzing these twelve treated countries and their donor pools, we construct control countries. We build a synthetic control unit for each treated country by weighting the donors so as to resemble each treated country as closely as possible in the time interval before the treatment. In the next two Sections we first describe how we extend the method to several variables after

which we present the results of this optimization.

Before we proceed with the analysis and the results we should mention that two countries in our sample experienced severe wars, namely Rwanda and Uganda. According to our data Rwanda experienced a drop in GDP of 60% in 1994. Moreover, the death rate almost doubled from 20 deaths per 1,000 inhabitants in 1987 to 38 deaths in 1993, while life expectancy dropped by 20 years. In Uganda the consequences were slightly less severe with a drop in GDP by 20% in 1979. The demographic variables show no clear peak for Uganda in terms of the death rate, even though it is generally quite high. For instance the death rate only dropped below 15 deaths per 1,000 inhabitants in 2004, while Kenya, the neighboring country shows a similar HIV prevalence but a death rate smaller than 10 deaths in the years 1987-1994. Given these wars the results for Rwanda and Uganda should be interpreted with caution, since we can not distinguish between wars and consequences of HIV.

3.5 Adapting the synthetic control group to multiple outcomes

Before we turn to the results we discuss more in detail how we extend the synthetic control group method to several variables. We form a vector consisting of all three variables with their pre- and post-treatment observations. Minimization then occurs with reference to this vector. Moreover, as we will describe below the method allows preferential weighting of the variables.

An important issue is how to weight the three variables respectively in order to obtain the minimum over all three dimensions. This minimum will not be unique, instead there will be a so called Pareto-frontier.[19] Similar to the Pareto-efficiency concept in the economic science, a Pareto frontier is defined as a point, where improvements with respect to one variable, can only be reached at losses in at least one other variable. Adopting this to our situation obtaining a lower RMSPE along the dimension of life expectancy will lead to a higher RMSPE along the dimensions of the death and/or birth rates.

Our minimization occurs along four steps. Firstly, we will divide each of the outcome variables by their individual one-dimensional minimum, i.e. the RMSPE that will be obtained by optimizing just with respect to this variable as in the standard case. In this way we compare the obtained minimum with the potential minimum, and therefore align the variables in relation to their absolute minimum. We perform this step in order to bring the three variables

[19]This is a well known concept mainly used in multi-objective optimization. There is a large literature in computational sciences on similar problems. Branke et al. (2008) contains both introductory material and recent algorithms for finding the frontier and obtaining desired solutions.

on a common scale. Without this step differences in life expectancy would receive the largest punishment, since the level of life expectancy is the largest and thus life expectancy will get the highest weight. Therefore this step is needed to ensure cross-variable comparability of the three considered outcomes.

As a second step we take the natural logarithm of the obtained values. In this way percentage deviations are minimized instead of levels. This is desirable since some of the variables either grow or decrease considerably over time, and in this way we ensure that a deviation in the first period where the level is comparably low (high) will get the same weight as a later deviation with a relatively high (low) value. This step will lead to a better cross-time or within comparability for all three variables.

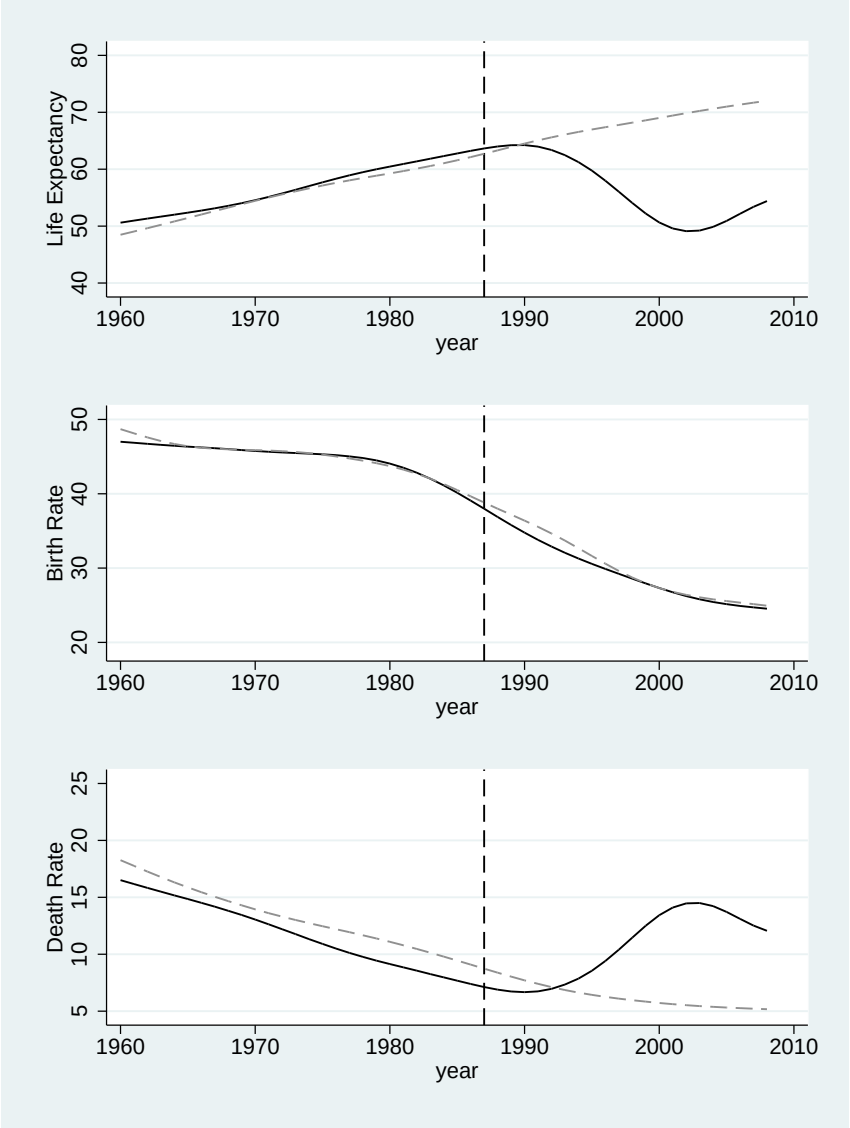
In the third step we specify a weight between the variables. As mentioned above there are many possible solutions along the Pareto-Frontier. Weighting helps to obtain solutions which are desired by the researcher. In our case we expect HIV to lead to very clear consequences on the death rate and life expectancy even if precision is low. For birth rates on the other hand we need a very high precision to verify whether HIV has an effect. Therefore, it is preferable to have a synthetic control group which is especially close in terms of the weight given to birth rates, while we will have a lower penalty for deviations along the other dimensions. In order to obtain this we multiply the birth rate by a certain factor – which is set at ten – to obtain a very precise estimate. The choice of this factor is quite arbitrary, however it should exceed one to ensure that the birth rate gets a higher weight as compared to the other two variables. We tried several different weights and found no noticeable difference between the estimates, which indicates that our results are quite robust to changes of this factor.

After these three steps we will obtain a vector Y consisting of the three outcome variables, adjusted by the weights we just described. This vector will be obtained for every country in the sample, both treated and non-treated. We then use the synthetic control group method suggested by Abadie and Gardeazabal (2003) described in Section 3.3 which weights the vectors of the control or donor countries in order to replicate the vector of the treated countries along all considered dimensions.

As Figure 3.1 shows for Botswana (and also the other countries as seen below) this will give us a very close match for the birth rate and a still relatively close matches for the death rate and life expectancy.

As seen in Table 2.4 Botswana's treatment occurred in 1987. This means we have 27 years of pre-treatment and 22 years post-treatment observations. Figure 3.1 displays the results of the minimization for Botswana. The dashed lines show the synthetic control group while the solid lines show the actual

Figure 3.1: Pre- and Post-Treatment in Botswana for all Three Considered Variables



outcomes along all three dimensions for Botswana. The graph shows that we achieve a synthetic control group with almost equal outcomes in terms of life expectancy and the birth rate before treatment, while there remains some difference in terms of the death rate. However, there is still a common trend between treatment and synthetic control before treatment and the treatment effect is still much larger than the apparent differences before treatment. For the other 11 countries in the sample we will obtain quite similar results. In fact Botswana's RMSPE along all three dimensions is quite close to the average RMSPE along the each of the three dimensions.

In the next Section we will apply this method to all twelve countries and try to answer the question how HIV influences life expectancy, death and birth rates.

3.6 Results

In this Section, we present the results of the synthetic control group estimates with the vector formed by combining the three variables. We first look at the weights in order to see which countries of the donor pool build the synthetic control group for each treated country. Afterwards we state the results of the three different outcome variables.

Looking at the donor pool, Table 3.2 displays the actual weights given to the different countries. As mentioned in Section 4 our donor pool was restricted to countries that are comparable with respect to life expectancy, the birth and death rate. This helps us to get a synthetic control group that is comparable also in these aspects.

Table 3.2: Country Weights

Donor	Botswana	Kenya	Lesotho	Malawi	Mozambique	Namibia	Rwanda	S. Africa	Swaziland	Uganda	Zambia	Zimbabwe	Mean Weight
Niger	-	0.451	0	0.8	0.596	-	0.592	0.003	0.363	0.344	0.399	-	0.394
Nicaragua	0.557	0.412	0	0	0	0.319	0	0	0.204	0.18	0.082	0.558	0.193
Pakistan	0.164	0.043	0.181	0	0	0.527	0	0	0.282	0.121	0.256	0	0.131
Comoros	0.025	0	0.124	0	0	0	0.408	0	0.001	0.27	0	0.294	0.094
Senegal	-	0	0.332	0.069	0	0	0	0	0	0	0	0.123	0.048
Algeria	0.192	0	0.114	0.131	0	0	0	0	0.011	0	0.09	0.025	0.047
Paraguay	0.063	0.095	0	-	-	0	-	0	0	0.085	0.041	0	0.032
Kazakhstan	-	-	0	-	-	-	-	0.085	0	-	-	-	0.028
Bangladesh	0	0	0	0	0.107	0	0	0.213	0	0	0	0	0.027
Nepal	0	0	0	0	0.224	0	0	0.001	0	0	0	0	0.019
Argentina	-	0	0.105	-	0	0.022	0	0.045	0	0	0	0	0.017
Austria	-	0	-	-	0.045	-	-	0	-	-	-	-	0.015
Iran	0	0	0	0	0	0.04	0	0	0.138	0	0	0	0.015
Mongolia	0	0	0	-	0	0.09	0	0.072	0	0	0	0	0.015
Fiji	0	0	0	-	-	0	-	0.131	0	0	0	0	0.015
Mexico	0	0	0	-	0	0	-	0	0	0	0.132	0	0.013
Bolivia	0	0	0	0	0	0	0	0.158	0	0	0	0	0.013
Qatar	0	0	0.018	-	0	0	-	0.097	0	0	0	0	0.012
...													
Sum	1	1	1	1	1	1	1	1	1	1	1	1	

This table displays how the synthetic control groups are formed for the different countries. It displays the weights W given to the most important donors (see equation 3.3 for more information. Donors with an average weight of less than 1% are not displayed. The "-" sign indicates that a certain country was not considered as potential donor because it was not close enough in terms of outcome variables. The note of the previous table and the text before it explains which countries are considered as donors.

Niger shows the highest average weight. Other important donors are Nicaragua and Pakistan. Moreover, there are some other noteworthy donors from African

countries such as Comoros, Senegal and Algeria. Next we turn to the three outcome variables to see how HIV prevalence affected each of them.

3.6.1 Life Expectancy

The first variable of interest is life expectancy. Before we turn to the result we consider at the fit in the pre-treatment period. This is displayed in Table 3.3 below.

For each country we display the country's actual values in the first column. In the second column we present the outcomes from the synthetic control group based on our vector outcome variable. Moreover, we compare this result to the outcomes of three other synthetic control groups: The third column shows the values if the minimization occurs with respect to one variable only, namely life expectancy. Moreover, columns four and five display the synthetic control group if minimization occurs only along the dimension of the birth and death rate respectively.

As expected if we minimize with respect to life expectancy only, we get the lowest RMSPE for most treated countries. Moreover, we can observe that the death rate and life expectancy are highly correlated, since for most cases we get a much lower RMSPE when we minimize with respect to the death rate as opposed to the birth rate.

Finally, turning to the RMSPE formed by minimizing over all three dimensions (2nd column), the obtained RMSPE is far below 10% in most cases.[20] A notable exception is Kenya with an RMSPE of 6.4 which measures around 14% of average life expectancy in Kenya. Here we fail to find a suitable control group along all three dimensions. The numbers further reveal that the birth rate seems to be the problem in terms of finding a suitable control group, since we obtain quite low RMSPEs when we minimize only along the dimension of life expectancy or the death rate. Also for Swaziland, Mozambique and Zimabwe the RMSPE is relatively high and again the birth rate seems problematic.

Summing up, we see that our suggested method of minimizing along three dimensions produces fairly good control groups in terms of life expectancy. However, the samples are clearly limited by the donors. The best example for this is Kenya, where it seems impossible to find a donor which is close along all three dimensions.

[20] Abadie et al. (2010) in their analysis obtain a value of around 3%.

Table 3.3: Means of Control and Treatment Groups for Life Expectancy

	Botswana					**Kenya**					**Lesotho**					**Malawi**				
	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death
Life Exp. (1960)	50.611	48.201	50.554	44.584	52.217	46.318	43.645	46.486	40.301	46.42	46.581	46.388	45.553	46.395	47.287	37.795	38.846	38.093	38.355	38.906
Life Exp. (Mid Year1)	53.567	52.923	53.768	48.091	55.682	51.023	46.147	51.07	42.012	50.789	48.805	49.459	49.39	50.162	49.93	39.74	39.819	40.166	39.392	39.975
Life Exp. (Mid Year2)	58.93	57.661	58.706	51.916	61.016	56.042	48.772	56.036	45.727	56.215	53.785	54.008	53.899	54.027	54.368	43.999	41.682	43.788	41.037	42.652
Life Exp. (T_0)	63.227	61.723	63.202	56.483	66.068	59.499	51.233	60.096	50.519	60.536	59.308	58.783	58.74	58.67	59.884	47.484	44.056	47.58	43.53	45.82
RMSPE	.	1.385	.128	6.336	2.211	.	6.465	.222	9.231	.398	.	.531	.584	.816	.832	.	2.04	.256	2.451	1.135
	Mozambique					**Namibia**					**Rwanda**					**South Africa**				
	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death
Life Exp. (1960)	35.019	38.17	37.799	38.982	38.553	46.885	48.653	46.842	47.265	47.702	42.283	39.846	42.138	39.998	41.938	49.048	49.168	48.838	48.878	48.616
Life Exp. (Mid Year1)	38.787	40.723	39.192	41.819	40.417	52.017	53.683	51.927	51.07	52.119	44.011	41.13	43.817	41.398	43.25	52.467	52.855	52.319	52.377	53.101
Life Exp. (Mid Year2)	42.712	44.56	41.567	45.903	42.947	57.342	57.864	57.318	54.767	57.235	44.671	42.722	45.025	43.142	44.594	56.792	58.243	56.875	56.858	57.999
Life Exp. (T_0)	43.142	48.913	44.674	50.681	45.247	61.778	61.695	61.779	58.354	61.811	46.666	44.611	46.666	45.153	46.931	61.375	62.962	61.509	61.418	62.47
RMSPE	.	3.123	1.241	4.372	1.784	.	1.188	.068	2.215	.293	.	2.348	.239	2.024	.492	.	1.05	.158	.089	.929
	Swaziland					**Uganda**					**Zambia**					**Zimbabwe**				
	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death	Real	M. Opt.	S. Life	S. Birth	S.Death
Life Exp. (1960)	44.258	43.829	44.085	43.753	45.216	43.965	43.789	43.804	41.947	45.151	45.095	45.068	45.001	41.788	46.944	51.497	45.13	51.343	41.757	52.958
Life Exp. (Mid Year1)	47.459	46.954	47.579	46.785	48.99	47.31	45.598	47.236	43.379	49.043	47.663	47.129	47.859	44.003	49.186	53.898	48.826	53.85	44.505	55.382
Life Exp. (Mid Year2)	54.331	50.211	54.386	50.159	55.486	50.979	48.105	49.774	45.613	51.531	51.054	49.4	50.599	46.885	52.035	56.987	53.646	57.098	47.902	58.585
Life Exp. (T_0)	60.533	53.677	60.445	53.557	62.275	49.979	49.985	51.2	47.778	52.482	52.082	51.531	52.831	50.647	54.443	60.935	57.188	60.569	52.068	61.912
RMSPE	.	3.885	.11	3.933	1.323	.	1.943	.81	4.072	1.565	.	1.031	.362	3.482	1.588	.	4.507	.146	9.252	1.419

The table compares the actual outcomes for each country with four different synthetic control groups. The first control group (M. Opt.) is formed by simultaneously minimizing the distance to the treated country in terms of life expectancy, birth and death rate. The three subsequent synthetic controls are formed by minimizing with respect to one variable only. The values are displayed for the year 1960, the year before the treatment year and the two years forming the first and second third of the distance between the the treatment and 1960 – Mid Year 1 and 2 respectively. The RMSPE in the last row is defined in equation (2.4).

Figure 3.2: Life Expectancy

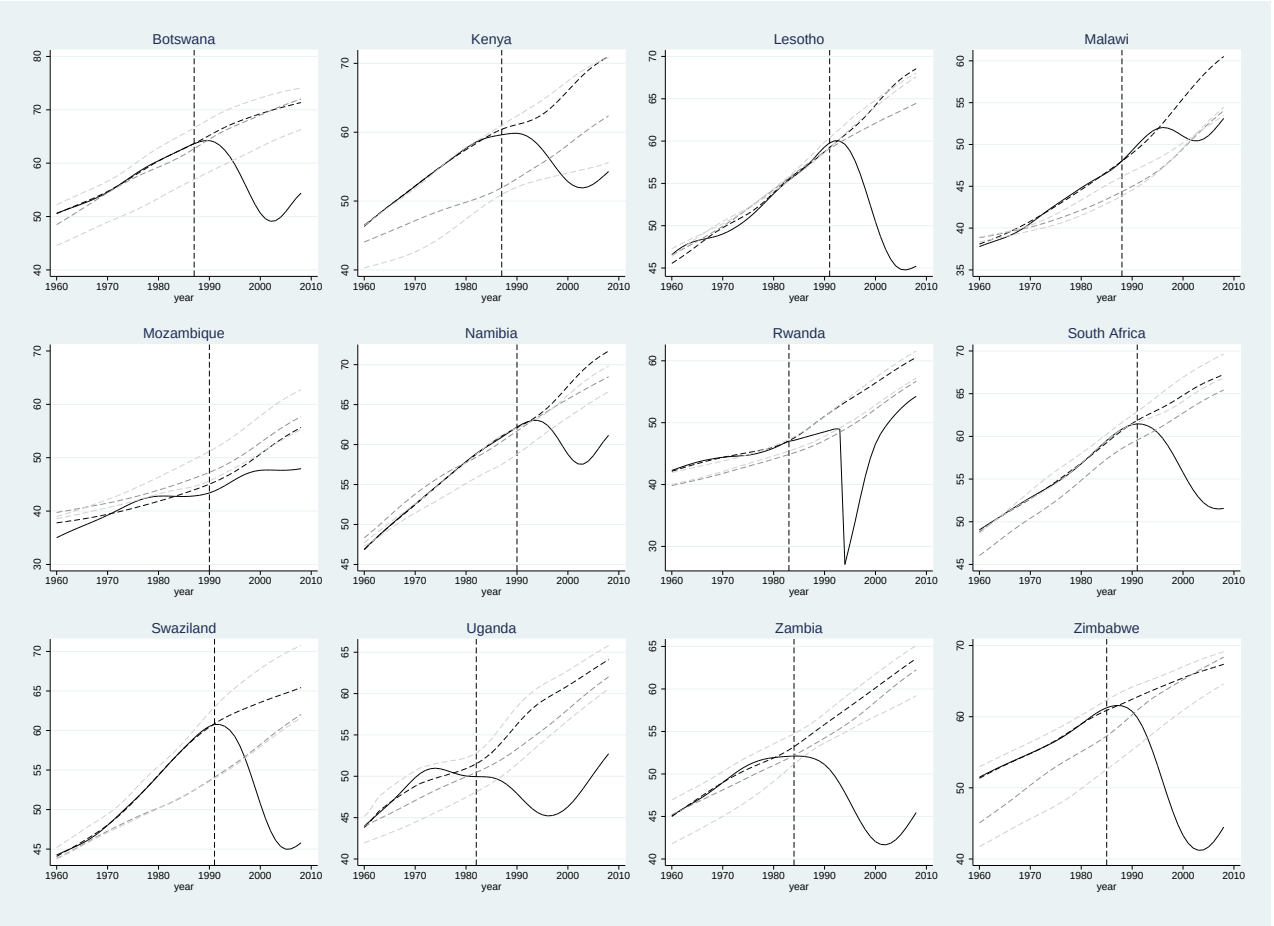


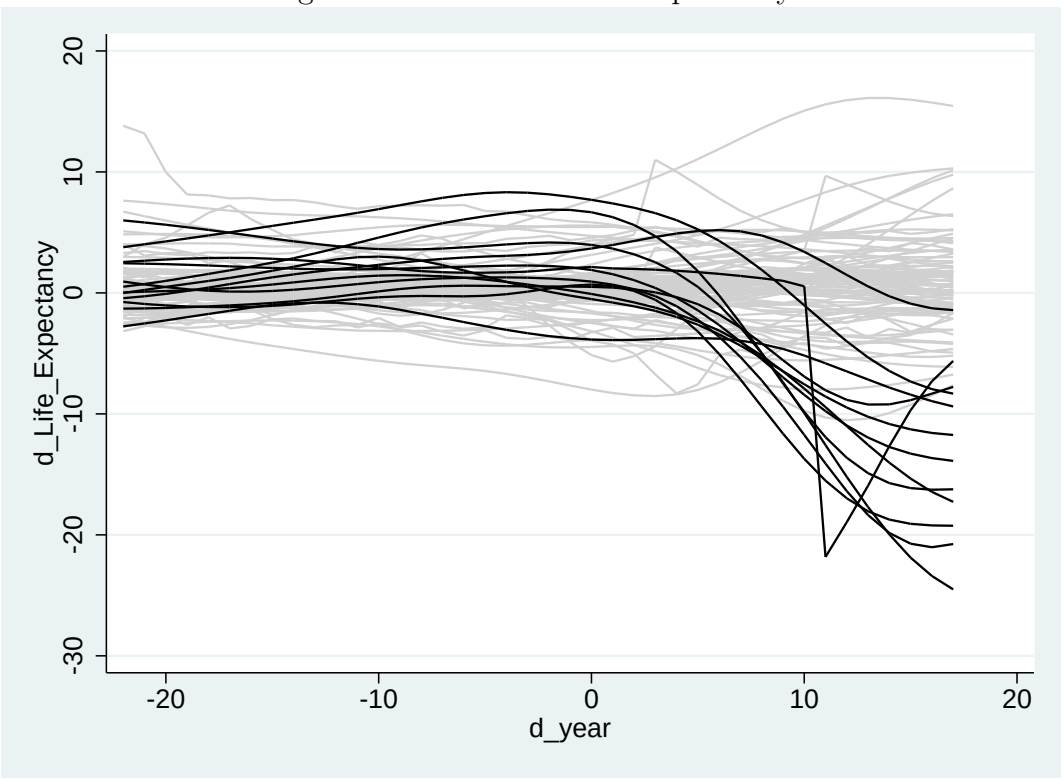
Figure 3.2 shows a graphical representation of the table just presented. For each country the black solid line stands for the actual value obtained, while dashed lines stand for the synthetic control groups. The black dashed line is usually closest to the real outcome since it shows the synthetic control group when we minimize only the distance in terms of life expectancy. The dark gray line presents the outcome when we minimize along all three dimensions. Finally, the two light gray lines show the synthetic control group when we only minimize the distance in terms of birth and death rate respectively. Again we can see that in most figures one of the light gray lines is fairly close to the real data, due to the high correlation of life expectancy and death rate. Moreover, we can observe that although the performance of our synthetic control group is weaker than minimizing only along the dimension of life expectancy, it is still quite close to the actual value.

This Figure shows both the fit in the pre-treatment interval and the result post-treatment. The treatment year is indicated by the vertical dashed line. In all countries we observe a negative treatment effect, which is expected, since higher HIV prevalence is bound to decrease life expectancy. Comparing the post-treatment outcomes of the different synthetic control groups we find that they are quite similar as long as there is a good fit before treatment. However, it seems that our minimization leads to a quite conservative estimate of the treatment effect: For almost all graph the treatment effect is largest when we minimize along the dimension of life expectancy only (black dashed line). Finally, we observe for most countries an increase in life expectancy towards the end of the observed time interval. Also for mortality (below) we observe a decreasing pattern in the last years. This is probably due to antiretroviral drugs which are used to reduce the health burden of HIV.

In order to analyze the treatment effects in more detail we perform placebo estimates. Here we randomly assign the treatment years of the 12 treated countries to the countries of the donor pool and perform the same minimization. Therefore the placebo estimates are created by minimizing the distance along all three dimensions and so we obtain a synthetic control for each country in the donor pool as well. Since the countries in the donor pool did not experience high HIV rates we should not observe a treatment effect there. By comparison of the results of the placebos and the treated countries we are able to interpret the previously measured effects. The results of this placebo estimates are displayed in Figure 3.3 and Table 3.4 below.

In Figure 3.3 we see the outcomes of the placebo estimates (gray lines), compared to the outcomes for the treated countries (black lines). The lines show the difference between treatment and synthetic control. The lines are aligned so that zero (on the x-coordinate) forms the treatment year for all

Figure 3.3: Placebos: Life Expectancy



countries. We see that before treatment all lines are quite close to zero (on the y-coordinate) and thus treatment and control group show similar trends. After treatment however, life expectancy decreases for almost all treated countries. The graph also displays that our estimate is someway conservative: Most treated countries show a positive deviation in life expectancy before treatment, and thus the synthetic control group tends to underestimate actual life expectancy before treatment. Since the treatment effect is negative this leads to a more conservative estimate of the treatment effect.

Table 3.4 further describes the placebo estimates and shows the treatment effects. Panel A of Table 3.4 displays the prediction errors before and after treatment for placebos and treated countries. Following Abadie et al. (2010) we calculate the prediction error in the pre-treatment period (1st column) and for the period after the treatment (2nd column) and build the ratio of the two RMSPEs (last column). The difference between the RMSPE and the Mean Prediction Error (4th column) is that for the MPE we don't square the error (and thus also don't take the square root of the mean afterwards). This way we can still interpret the sign, however only the RMSPE are directly comparable before and after the treatment. Finally in the fifth column we report the difference between synthetic control group and treatment country in the final period (2008). Since the deviation is cumulative also this variable builds an important indicator.

Table 3.4: Prediction Errors: Placebo Analysis Life Expectancy

PANEL A: Data

Rank	Country	Treated	RMSPE	RMSPEpost	Est. Treatment Effect	Final Period	RMSPE ratio
1	Lesotho	1	0.644	12.500	-10.497	-19.247	19.401
2	Zambia	1	0.883	12.225	-10.556	-16.783	13.838
3	Botswana	1	1.071	13.891	-11.727	-17.626	12.967
4	Lao PDR	0	0.435	5.549	4.936	6.521	12.747
5	Guatemala	0	0.609	7.207	6.457	8.881	11.826
6	Fiji	0	0.375	3.237	-3.102	-4.496	8.641
7	Nicaragua	0	0.944	7.294	6.294	11.065	7.723
8	Georgia	0	0.280	2.086	-1.893	-3.703	7.452
9	India	0	0.602	4.358	-4.223	-6.112	7.242
10	Iran Islamic Rep	0	0.797	5.123	4.975	6.004	6.429
..							
12	Namibia	1	0.951	6.000	-4.967	-7.338	6.307
..							
19	Uganda	1	1.817	8.588	-7.886	-9.296	4.726
..							
26	South Africa	1	2.346	8.362	-6.625	-13.883	3.565
..							
27	Zimbabwe	1	4.720	16.736	-13.041	-23.881	3.546
..							
32	Rwanda	1	2.357	7.705	-4.091	-2.440	3.269
..							
38	Swaziland	1	3.766	10.332	-6.351	-16.234	2.743
..							
51	Mozambique	1	2.727	6.097	-5.814	-9.758	2.236
..							
60	Malawi	1	1.956	3.387	2.134	-0.932	1.731
..							
78	Kenya	1	6.065	6.252	-1.621	-8.080	1.031
..							

PANEL B: Regression

	Est. Treatment Effect	Final Period	RMSPE ratio
Africa	-0.00873	-0.180	-0.236
	(0.643)	(0.885)	(0.765)
Treated	-8.497***	-14.59***	4.306***
	(1.245)	(1.714)	(1.171)
Constant	0.495*	0.600	2.667***
	(0.295)	(0.406)	(0.359)
N	101	101	101
R^2	0.364	0.475	0.147

Panel A shows the estimated effects for each country. We distinguish treated countries (HIV prevalence≥10%) and placebo countries (HIV precalence≤1%). The RMSPE is defined in equation (2.4). We calculate it before the intervention (column 3), afterwards (column 4) and their ratio (column 7) by which the countries are ranked. Column 5 equals the RMSPE post, however without squaring the difference and taking the square root of the mean. In column 6 we show the difference of the outcome variable in the final period (2008). In Panel B we present the results from the regression $\Pi = \alpha + \beta_1 Treated + \beta_2 Africa$ for the treatment and African dummy, where Π represents the column title. Thus each column represents one regression result. For the estimated treatment effect and the effect in the final period the observations are weighted by the inverse of their RMSPE. The belligerent countries Rwanda and Uganda are excluded from this regression.

Countries in Panel A are ordered based on their RMSPE ratio. The table shows that the first three countries in the list – Lesotho, Zambia and Botswana – are treated and show treatment effects of around 10 years. In fact as Table 2.2 reveals Botswana and Lesotho have a very high prevalence rate of above 20%, therefore it is only expected that the treatment effect is largest for them. Also Swaziland has a similar high prevalence rate; however, the fit before treatment is quite low and therefore the estimated treatment effect is only around 6 years. In order to analyze the average treatment effect for countries having a prevalence rate above 10% we turn to Panel B of Table 3.4.

In Panel B we present the regression results of the following equation:

$$\Pi = \beta_0 + \beta_1 Treated + \beta_2 Africa \tag{3.5}$$

The displayed result represents the β_i coefficients. Moreover, Π represents the outcome variable and thus varies from column to column. The aim of this regression is to find out whether treated and African countries are significantly different from the placebo countries with respect to the treatment effect and the RMSPE ratio. Panel A includes two countries with severe wars (Rwanda and Uganda), which might contaminate our treatment effect, therefore the regressions are estimated without these two countries. Since these two countries are not exceptional in terms of HIV prevalence and we are estimating an average effect, this should not affect the results too much. [21]

The estimated coefficients reveal that the affected countries have on average a life expectancy which is 8.5 years below their synthetic controls. Moreover, in the final year the difference is 14.5 years on average. Furthermore, the RMSPE ratio is more than four points higher than for the other countries. Therefore we can conclude that HIV had quite large effects on life expectancy in the countries with the highest prevalence rates. Finally, the African dummy is small and insignificant in all specifications. Next we analyze the birth and death rates presenting the same tables and figures already shown for life expectancy.

3.6.2 Birth Rate

Also for the birth rate we start by comparing the RMSPE of the different synthetic control groups in Table 3.5. Here the advantage of our estimator becomes most obvious, since minimization along the dimension of life expectancy or death rate only leads to very large RMSPEs. Of course it helps that we put a higher weight on this outcome variable as compared to others – as described at the end of the previous Section – however as just seen for life expectancy, the side effects are quite low. The RMSPE is below one for all countries except

[21] In fact including the two countries leads to a slightly different point estimates, but for all variables the estimates do not differ by more than one standard error.

for Malawi and Kenya, and is smaller than 3% for all countries. Moreover, minimizing only the distance in terms of life expectancy or birth rates results in much higher RMSPEs which exceed 5 in almost all cases.

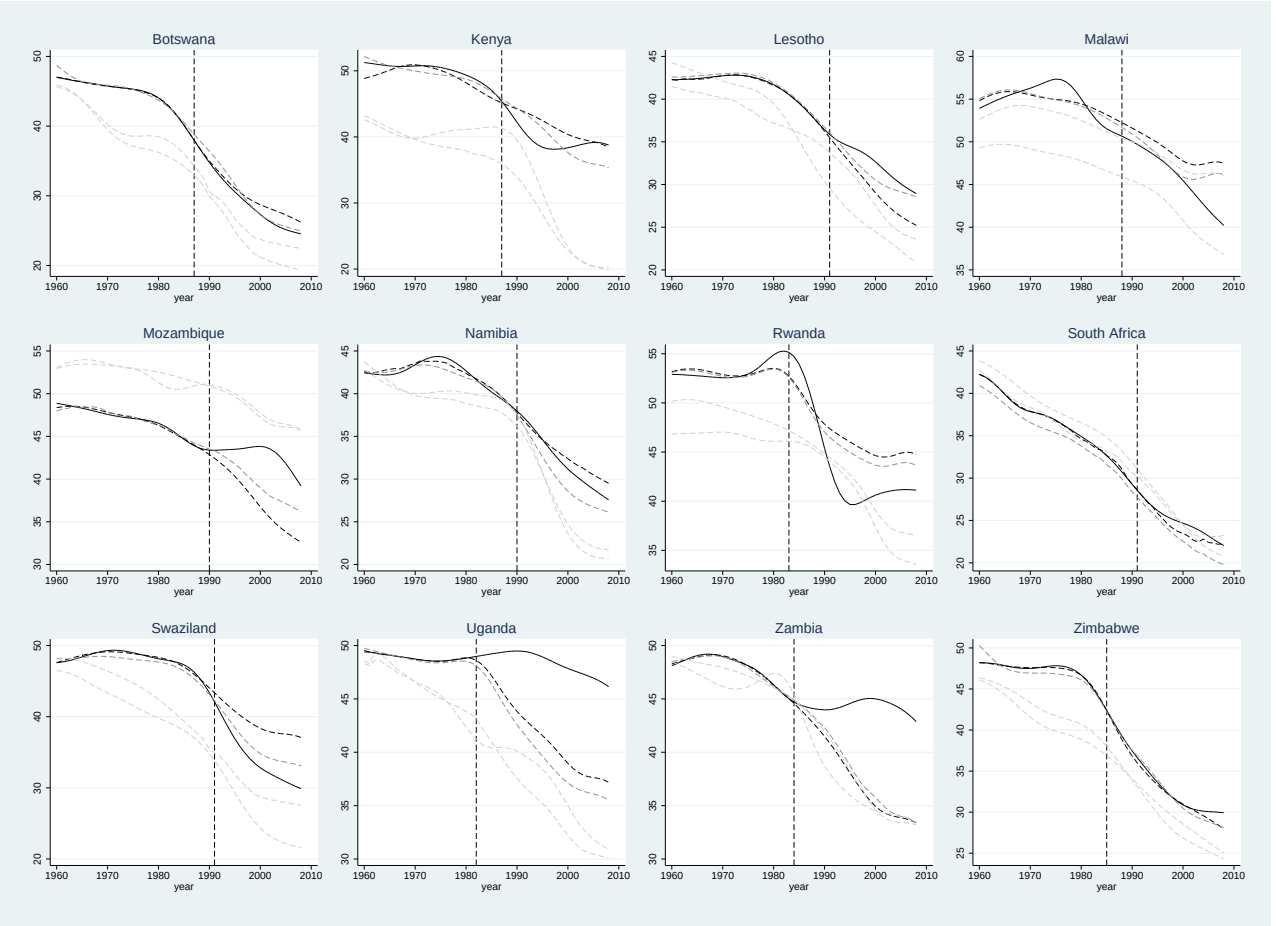
Figure 3.4 confirms this finding. Before treatment the actual outcomes and the synthetic control group resulting from our optimization are very close for most countries. Moreover, the two light gray lines are quite far away from the actual birth rate. Summing up, our control group seems to be very close to our treatment country before treatment. After treatment however, the results are mixed and seem to be quite heterogeneous. For Botswana and Lesotho, which had very large treatment effects in terms of life expectancy, the birth rate seems unaffected by the high HIV prevalence. For other countries, like Zambia and Mozambique we see that the birth rate in the synthetic control decreases, while it keeps rather constant in the treated countries. This together with the evidence on life expectancy observed above suggests a mechanism similar to the one suggested by Chen (2010). While the synthetic control group experienced an increase in life expectancy and the subsequent decrease in fertility, this does not seem to have occurred for some treated countries. However, as seen in the figure this does not seem to be the case for most other treated countries.

Table 3.5: Means of Control and Treatment Groups for Birth Rate

	Botswana					**Kenya**					**Lesotho**					**Malawi**				
	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death
Birth Rate (1960)	47	48.343	47.031	45.621	45.881	51.26	52.39	48.84	42.603	43.222	42.26	42.566	42.321	44.267	41.446	53.92	55.037	54.82	49.297	52.638
Birth Rate (Mid Year1)	46.01	45.67	45.926	41.68	41.197	50.67	50.256	50.81	39.911	40.249	42.68	42.938	42.65	42.317	40.268	55.9	55.947	55.757	49.479	54.269
Birth Rate (Mid Year2)	45.03	44.459	44.914	38.596	36.79	50.14	49.194	49.45	40.935	38.341	41.69	41.934	41.626	39.477	37.201	56.56	54.557	54.741	48.151	52.964
Birth Rate (T_0)	39.08	39.405	39.091	35.206	33.646	46.3	46.108	45.51	41.489	36.363	36.42	36.654	36.252	30.345	34.259	50.93	52.045	52.554	46.122	50.665
RMSPE	.	.483	.061	4.844	6.111	.	.73	1.049	9.071	10.622	.	.254	.091	3.014	3.067	.	1.316	1.355	6.609	2.227
	Mozambique					**Namibia**					**Rwanda**					**South Africa**				
	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death
Birth Rate (1960)	48.88	48.264	48.359	53.048	52.939	42.44	42.996	42.57	42.807	43.732	52.9	53.095	53.169	50.157	46.84	42.27	42.512	42.224	42.737	43.851
Birth Rate (Mid Year1)	47.72	48.058	47.97	53.684	53.372	43.04	43.392	43.368	40.127	40.005	52.67	53.054	53.241	50.036	46.975	38.03	38.408	38.114	38.56	40.159
Birth Rate (Mid Year2)	46.73	46.501	46.537	52.642	51.723	43.17	42.245	42.679	40.205	39.046	52.76	52.631	52.777	48.979	46.675	34.92	35.069	34.624	34.377	36.536
Birth Rate (T_0)	43.47	43.811	43.267	51.069	51.107	38.48	38.395	38.449	38.334	36.77	55.28	53.084	53.175	47.432	46.124	29.28	29.471	29.251	30.643	31.823
RMSPE	.	.27	.221	6.146	5.658	.	.606	.406	2.425	3.069	.	.925	.914	4.53	6.786	.	.264	.147	.647	1.908
	Swaziland					**Uganda**					**Zambia**					**Zimbabwe**				
	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death	Real	M. Opt.	S. Birth	S. Life	S.Death
Birth Rate (1960)	47.59	48.088	47.592	48.253	46.469	49.53	49.386	49.401	48.289	48.558	48.11	48.629	48.281	48.974	48.465	48.18	49.808	48.235	46.116	46.408
Birth Rate (Mid Year1)	49.19	48.477	49.087	46.705	43.695	49.03	48.702	49.033	47.628	47.978	49.2	48.974	49.103	48.246	46.845	47.61	46.8	47.681	43.23	44.566
Birth Rate (Mid Year2)	48.14	47.666	48.341	42.448	39.792	48.54	48.043	48.532	45.668	45.302	48.12	47.939	48.094	47.13	46.127	47.81	46.398	47.523	39.677	41.535
Birth Rate (T_0)	43.37	43.134	44.047	35.42	34.811	48.9	48.046	48.807	41.757	43.571	45.02	45.238	45.028	45.233	46.329	43.45	42.842	43.54	37.389	38.713
RMSPE	.	.674	.313	5.159	7.029	.	.502	.088	3.61	3.096	.	.221	.091	.747	1.794	.	.979	.131	6.146	4.651

The table compares the actual outcomes for each country with four different synthetic control groups. The first control group (M. Opt.) is formed by simultaneously minimizing the distance to the treated country in terms of life expectancy, birth and death rate. The three subsequent synthetic controls are formed by minimizing with respect to one variable only. The values are displayed for the year 1960, the year before the treatment year and the two years forming the first and second third of the distance between the the treatment and 1960 – Mid Year 1 and 2 respectively. The RMSPE in the last row is defined in equation (2.4).

Figure 3.4: Birth Rates



In order to see whether the effects are significant we compare the outcomes to the placebos in Figure 3.5.

Figure 3.5: Placebos: Birth Rates

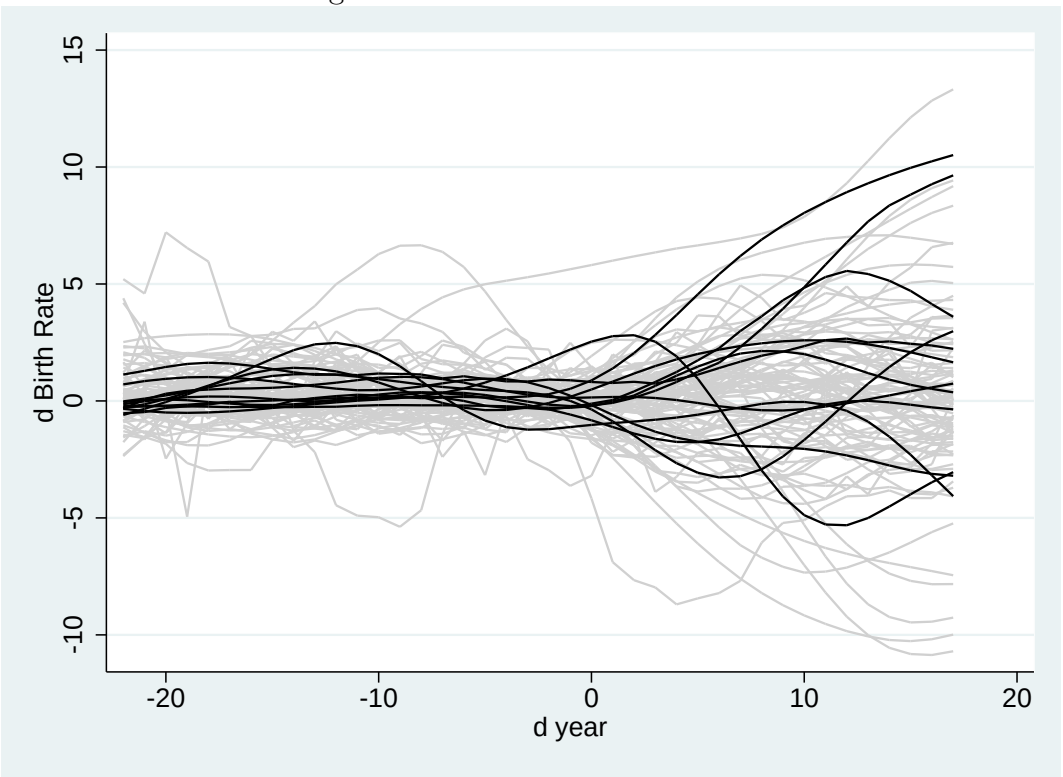


Figure 3.5 reveals that for most countries both in the treatment group (black lines) and the placebos (gray lines) we find fairly good controls. After treatment however the estimates largely diverge. Some treated countries seem to have quite large negative effects, while others exhibit positive effects. This confirms the effect heterogeneity already observed in Figure 3.4.

Finally, Table 3.6 allows a better analysis of treatment effects and their significance for the countries. Panel B reveals that there is no overall effect that can be measured. Moreover, the explanatory power is generally quite low in these regressions. While the R^2 in the regressions above was close to 0.5, here it is only around 0.1. Therefore once more the results are mixed.

Looking at the effects in the single countries in Panel A we see a quite large positive effect of 8 additional births in Uganda. For Uganda however, we can not distinguish between the treatment effect due to HIV and the effect caused by the war – which is why it is excluded from the regressions together with Rwanda. Including Rwanda and Uganda leads to a slightly larger treatment effect, however all point estimates are within one standard error of the current estimate. The only country showing a treatment effect which is large, significant at a 3% level and can be attributed to HIV is Zambia indicating 6 additional births per 1,000 inhabitants. The effects seen for the other countries

Table 3.6: Prediction Errors: Placebo Analysis Birth Rate

PANEL A: Data

Rank	Country	Treated	RMSPE	RMSPEpost	Est. Treatment Effect	Final Period	RMSPE ratio
1	Uganda	1	0.188	8.580	8.197	10.651	45.691
2	Bhutan	0	0.222	7.952	-7.774	-9.989	35.793
3	Zambia	1	0.202	6.961	6.003	9.411	34.392
4	Madagascar	0	0.256	5.978	5.926	6.441	23.381
5	Guatemala	0	0.258	6.008	5.285	7.363	23.298
6	Somalia	0	0.394	7.793	6.888	12.141	19.765
7	Bangladesh	0	0.411	6.159	-5.966	-8.413	14.971
8	Iran Islamic Rep	0	0.553	7.887	-6.857	-9.559	14.256
9	Oman	0	0.380	4.643	-3.787	-7.827	12.225
10	Mozambique	1	0.319	3.630	3.282	2.978	11.363
..							
26	Lesotho	1	0.257	1.348	1.201	0.369	5.236
..							
33	Rwanda	1	0.782	3.285	-2.192	-2.506	4.200
34	Kenya	1	0.695	2.437	-0.060	3.405	3.505
..							
42	Swaziland	1	0.681	2.113	-2.059	-3.196	3.103
..							
45	Namibia	1	0.671	2.027	2.015	1.448	3.023
..							
60	Botswana	1	0.499	0.954	-0.740	-0.404	1.913
61	Malawi	1	1.326	2.489	-1.689	-6.002	1.878
..							
73	South Africa	1	1.217	1.918	1.836	2.229	1.576
..							
87	Zimbabwe	1	0.808	0.750	0.371	1.849	0.928
..							

PANEL B: Regression

	Est. Treatment Effect	Final Period	RMSPE ratio
Africa	0.409	1.060	2.502
	(0.753)	(1.073)	(1.557)
Treated	1.653	1.891	0.472
	(1.071)	(1.525)	(2.383)
Constant	0.00249	-0.192	3.717***
	(0.359)	(0.511)	(0.730)
N	101	101	101
R^2	0.049	0.053	0.039

Panel A shows the estimated effects for each country. We distinguish treated countries (HIV prevalence≥10%) and placebo countries (HIV precalence≤1%). The RMSPE is defined in equation (2.4). We calculate it before the intervention (column 3), afterwards (column 4) and their ratio (column 7) by which the countries are ranked. Column 5 equals the RMSPE post, however without squaring the difference and taking the square root of the mean. In column 6 we show the difference of the outcome variable in the final period (2008). In Panel B we present the results from the regression $\Pi = \alpha + \beta_1 Treated + \beta_2 Africa$ for the treatment and African dummy, where Π represents the column title. Thus each column represents one regression result. For the estimated treatment effect and the effect in the final period the observations are weighted by the inverse of their RMSPE. The belligerent countries Rwanda and Uganda are excluded from this regression.

are quite small and not significantly different from the placebos.

3.6.3 Death Rate

The last variable of interest is the death rate. Looking first at the RMSPE in Table 3.7 we see that it is quite high exceeding 10% for Botswana, Kenya, Mozambique, Swaziland and Zimbabwe. Again this is caused by the fact that we want to obtain a higher precision in terms of birth rate. Once more we can confirm that the death rate and life expectancy are highly correlated, since we find a quite low RMSPE in terms of the death rate, when we minimize in terms of death rate or life expectancy. This is also seen in Figure 3.6 where one light gray line is quite close to the actual outcome, while the second light gray line created by minimizing along the birth rate is quite far away from the real outcome. The synthetic control group obtained by minimizing along all three dimensions on the other hand is somewhere in between these two lines.

Once more we compare our outcomes with the placebo estimates in order to see the treatment effects and significance. While Figure 3.7 reveals that there is already some deviation before treatment for many countries, the treatment effects that can be observed afterwards are all positive and seem to be quite large, exceeding most of the placebo outcomes.

Table 3.7: Means of Control and Treatment Groups for Death Rate

	Botswana					**Kenya**					**Lesotho**					**Malawi**				
	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth
Death Rate (1960)	16.51	18.15	16.579	17.333	21.386	20.21	20.339	20.211	20.365	24.29	19.2	19.309	19.206	20.869	19.976	28.28	25.343	25.351	25.176	25.797
Death Rate (Mid Year1)	13.82	14.595	13.726	14.386	18.386	16.3	17.961	16.289	16.338	23.1	17.56	17.412	17.523	17.267	17.157	25.18	25.069	24.994	23.869	25.607
Death Rate (Mid Year2)	10.12	11.89	10.179	10.963	15.557	12.65	15.92	12.604	12.989	19.61	14.07	14.266	14.148	14.014	14.388	22.88	23.157	22.519	21.001	24.241
Death Rate (T_0)	7.39	9.056	7.484	8.616	11.814	10.12	13.874	9.864	9.815	15.2	10.67	10.802	10.669	10.41	10.95	18.71	20.189	19.648	18.186	22.068
RMSPE	.	1.453	.06	.844	4.882	.	2.783	.106	.238	6.278	.	.209	.124	.646	.393	.	1.23	1.084	1.632	1.943
	Mozambique					**Namibia**					**Rwanda**					**South Africa**				
	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth
Death Rate (1960)	28.7	25.021	25.733	25.821	24.717	18.58	18.606	18.606	19.513	20.018	22.23	24.461	21.849	22.52	24.526	17.36	16.981	17.446	18.026	17.973
Death Rate (Mid Year1)	24.97	23.115	24.552	25.3	22.604	14.92	14.907	14.927	15.221	17.237	20.81	23.362	21.115	21.604	23.629	14.25	14.24	14.265	15.011	15.23
Death Rate (Mid Year2)	21.5	19.754	22.357	23.389	19.423	11.56	12.009	11.535	11.761	14.407	20.28	21.754	20.223	20.639	22.235	10.84	10.906	10.818	11.927	12.077
Death Rate (T_0)	20.74	16.155	20.347	20.897	15.889	8.71	9.361	8.721	8.611	11.672	18.63	20.127	18.626	19.269	20.793	8.33	8.375	8.277	9.552	9.358
RMSPE	.	2.776	1.108	1.353	3.113	.	.445	.024	.367	2.572	.	1.974	.264	.559	2.318	.	.206	.048	1.055	1.131
	Swaziland					**Uganda**					**Zambia**					**Zimbabwe**				
	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth	Real	M. Opt.	S.Death	S. Life	S. Birth
Death Rate (1960)	20.18	21.33	20.613	21.774	22.417	20.71	20.472	20.592	21.786	22.48	19.46	19.928	19.59	20.606	22.861	15.61	20.167	15.596	16.498	23.2
Death Rate (Mid Year1)	17.98	18.476	17.678	18.962	20.381	17.99	18.829	17.86	19.252	21.413	17.77	18.179	17.593	18.641	21.338	13.88	16.834	13.918	14.5	20.715
Death Rate (Mid Year2)	13.08	15.565	13.194	13.789	17.68	15.36	16.728	15.987	17.217	19.728	15.31	16.075	15.5	16.736	18.895	11.85	13.726	11.817	12.181	17.966
Death Rate (T_0)	9.13	12.345	9.159	9.691	14.796	16.1	15.342	15.134	15.621	18.268	14.33	14.12	13.913	15.083	15.547	9.23	11.06	9.27	9.923	14.375
RMSPE	.	2.112	.206	.964	4.017	.	.984	.571	1.341	3.445	.	.516	.207	1.116	3.214	.	2.707	.03	.594	6.414

The table compares the actual outcomes for each country with four different synthetic control groups. The first control group (M. Opt.) is formed by simultaneously minimizing the distance to the treated country in terms of life expectancy, birth and death rate. The three subsequent synthetic controls are formed by minimizing with respect to one variable only. The values are displayed for the year 1960, the year before the treatment year and the two years forming the first and second third of the distance between the the treatment and 1960 – Mid Year 1 and 2 respectively. The RMSPE in the last row is defined in equation (2.4).

Figure 3.6: Death Rates

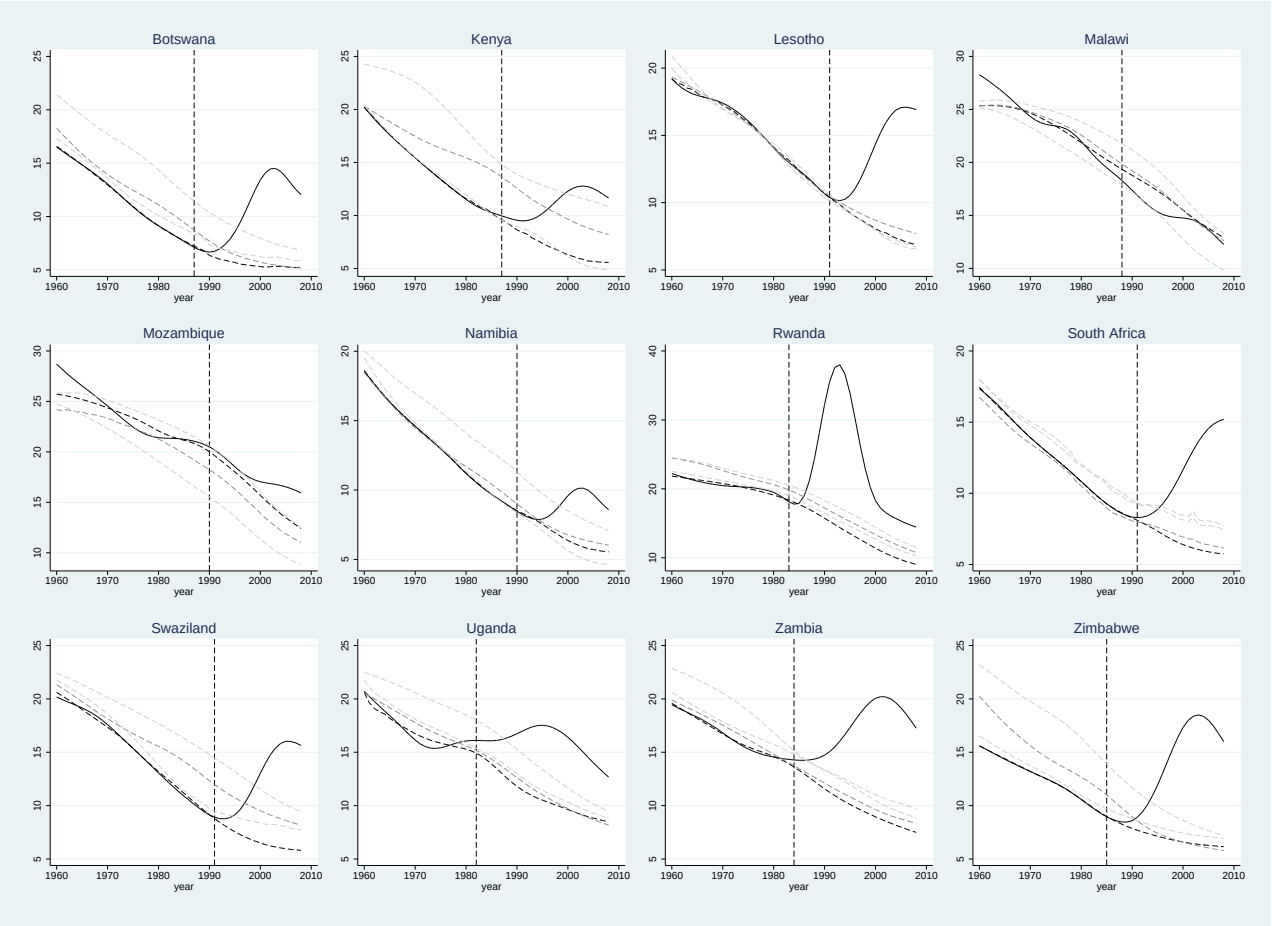


Figure 3.7: Placebos: Death Rates

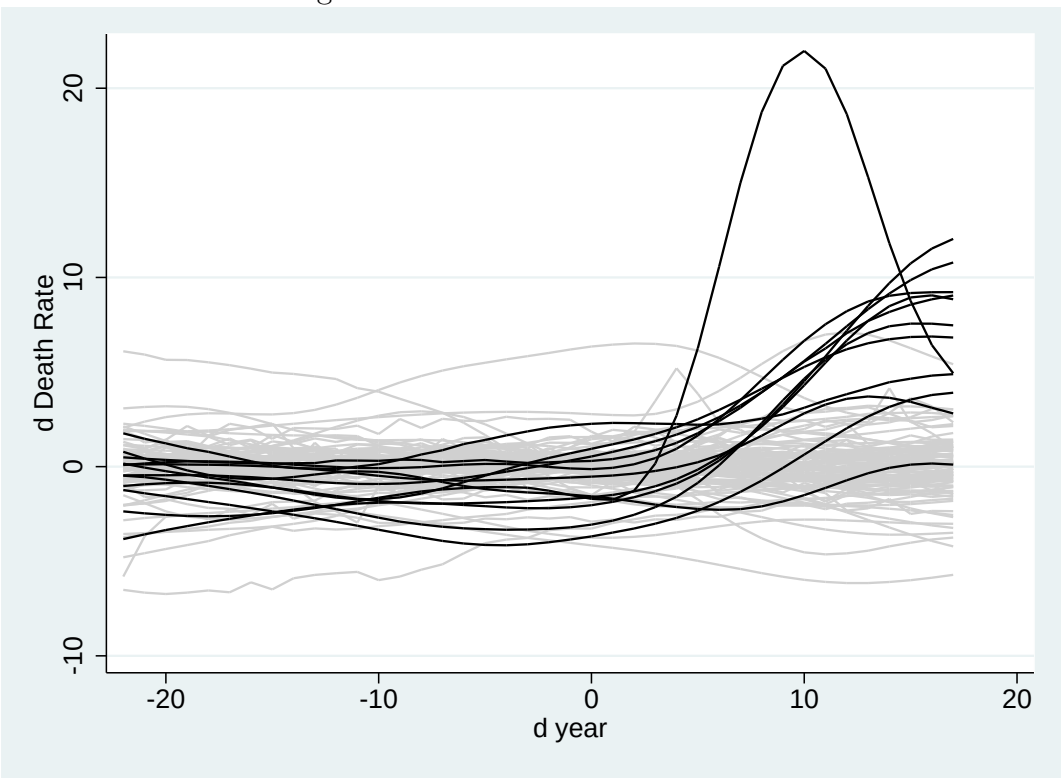


Table 3.8 confirms this and show that among the 10 countries with the highest RMSPE ratio 5 are treated countries (Panel A). The estimated treatment effect is around four additional deaths on average and 6.7 additional deaths in the final period. Finally, the RMSPE ratio is significantly higher in the treated countries.

3.7 Discussion

In this Chapter, we estimate the effects of the HIV pandemic on demographics in the twelve most heavily affected countries. We employ an extension of the synthetic control group approach (Abadie and Gardeazabal 2003), which allows the inclusion of several dependent variables. Using this method , we find suitable control groups based on key demographic indicators, namely life expectancy, death and birth rates.

The results reveal that HIV significantly decreased the life expectancy by almost 15 years, and the effects range from one year for Malawi to almost 24 years for Zimbabwe. As expected, life expectancy decreased more in the more heavily affected countries. In terms of death rate the average effect is close to seven deaths per 1,000 inhabitants. Again there is a wide range of effects, with Malawi showing no effect at all, while the largest effect is measured for Zimbabwe with slightly more than ten additional deaths due to HIV.

Table 3.8: Prediction Errors: Placebo Analysis Death Rate

PANEL A: Data

Rank	Country	Treated	RMSPE	RMSPEpost	Est. Treatment Effect	Final Period	RMSPE ratio
1	Lao PDR	0	0.109	3.451	-3.230	-3.652	31.666
2	Lesotho	1	0.216	6.038	5.127	9.223	27.983
3	Guatemala	0	0.103	1.802	-1.456	-2.336	17.570
4	Zambia	1	0.602	7.490	6.707	8.908	12.445
5	South Africa	1	0.445	5.470	4.721	9.046	12.290
6	Mauritania	0	0.257	1.637	1.544	2.463	6.358
7	Namibia	1	0.390	2.345	1.869	2.564	6.011
8	Bangladesh	0	0.502	2.934	-2.547	-4.402	5.844
9	Madagascar	0	0.252	1.439	1.276	0.152	5.716
10	Rwanda	1	1.985	10.911	8.438	3.704	5.497
..							
13	Uganda	1	1.027	5.081	4.847	4.495	4.945
..							
16	Botswana	1	1.507	5.728	4.393	6.887	3.800
..							
19	Zimbabwe	1	2.788	7.888	6.086	10.216	2.829
..							
27	Swaziland	1	2.066	4.838	2.993	7.469	2.341
..							
39	Mozambique	1	1.958	3.420	3.310	4.924	1.747
..							
57	Malawi	1	1.216	1.459	-1.116	-0.359	1.200
..							
65	Kenya	1	2.808	2.892	0.810	3.434	1.030
..							

PANEL B: Regression

	Est. Treatment Effect	Final Period	RMSPE ratio
Africa	0.504	0.668	-0.276
	(0.329)	(0.429)	(1.125)
Treated	3.831***	6.691***	5.318***
	(0.568)	(0.741)	(1.721)
Constant	-0.286*	-0.385*	2.125***
	(0.157)	(0.205)	(0.527)
N	101	101	101
R^2	0.421	0.552	0.109

Panel A shows the estimated effects for each country. We distinguish treated countries (HIV prevalence≥10%) and placebo countries (HIV precalence≤1%). The RMSPE is defined in equation (2.4). We calculate it before the intervention (column 3), afterwards (column 4) and their ratio (column 7) by which the countries are ranked. Column 5 equals the RMSPE post, however without squaring the difference and taking the square root of the mean. In column 6 we show the difference of the outcome variable in the final period (2008). In Panel B we present the results from the regression $\Pi = \alpha + \beta_1 Treated + \beta_2 Africa$ for the treatment and African dummy, where Π represents the column title. Thus each column represents one regression result. For the estimated treatment effect and the effect in the final period the observations are weighted by the inverse of their RMSPE. The belligerent countries Rwanda and Uganda are excluded from this regression.

Comparing the estimated effects with other studies the estimates are generally similar, even though their exist notable differences. The report by UN Population Division (2004) is the only study we are aware of, that estimates counterfactuals and actual rates of demographic variables for several countries. While they do not have estimates on the birth rates, they also cover the crude death rate and life expectancy. The counterfactual without HIV was formed by comparable countries and their death rates and causes (see Buettner et al. 2003, for more information on the methodology). In terms of the sample the paper distinguishes between the very heavily affected countries Botswana Lesotho, Namibia, South Africa, Swaziland, Zambia and Zimbabwe and other five heavily affected countries Cameroon, Central African Republic, Kenya, Malawi and Mozambique. Comparing this to our sample the estimated HIV prevalence they consider is higher in Cameroon and the Central African Republic, while our data sources consisting of Oster (2007) and UNAIDS (2010) estimate higher prevalence rates for Uganda and Rwanda. In terms of results the study indicates a decrease in life expectancy of 5.2 years in the years 2005-2010 for the seven very heavily affected countries, and a decrease of 19.7 years in the five heavily affected countries. These estimates are quite similar to the 14 years we estimate on average for the twelve most heavily affected countries. However contrary to their results, we estimate the largest effects for the most heavily affected countries. In terms of the crude death rate the UN Population Division (2004) estimates effects of 17.9 and 9.5 additional deaths for the very heavily affected countries and heavily affected countries respectively. Our estimate is more conservative and estimates only seven additional deaths on average. In summary both our study and the study of the UN Population Division (2004) suggest large enough effects on life expectancy and the death rate to induce changes in behavior, especially for the more heavily affected countries.

We also analyze these behavioral effects by looking at the birth rate. For the birth rate we obtain very precise estimates as well, indicating no overall effect, since the RMSPE after treatment is not significantly larger for the treated countries. Moreover, the estimated effects are both positive and negative and quite small in size. Zambia and Mozambique form a notable exception with nine additional births and three additional births respectively, per 1,000 inhabitants due to HIV. The data suggests that while the birth rate in the synthetic control group decreased, Zambia's and Mozambique's birth rate stayed rather constant. In terms of mechanisms the large decrease in terms of life expectancy in Zambia and Mozambique might be responsible for the non-decrease in terms of the birth rate, as suggested by Chen (2010). However, it remains an open issue why similar effects did not occur in other countries.

Kalemli-Ozcan (2012) estimates the effect of HIV on fertility in a recent study. The used sample consists of 44 countries and the author finds positive effects in between country comparisons, while results turn insignificant in within-country comparisons. Our results are in fact quite similar. Also we find positive effects for some countries, while on average the effect is insignificant for the twelve countries analyzed.

We estimate the consequences and employ placebo studies to see whether there are significant deviations from the estimated counterfactuals. Here we find that there are indeed large differences between the countries with high HIV prevalence rates and other countries with similar (initial) demographic conditions. Therefore it seems that the method is suitable to estimate the effects. However, we also reveal a drawback of the method. In Rwanda and Uganda severe wars occur during treatment and the method does not allow to disentangle treatment effects from HIV from other effects.

Before we turn to the impact of the Spanish Flu in Sweden, we briefly recapitulate what we learned in terms of effects of AIDS in Africa. The clearest finding is that the country with the largest loss in terms of GDP in Chapter 2, Zambia, was also the only country to exhibit a positive, large and significant effect on the birth rate. The large decrease in GDP can thus be explained by the fact that due to HIV its birth rate did not decrease as in other countries. Also for Lesotho we find similar effects, with a smaller decrease in birth rate as compared to other countries and GDP losses. However, for Lesotho the effects are less pronounced and not significant. Mozambique and Uganda show similar effects with respect to the birth rate. The results for Zambia and Lesotho would predict large effects in terms of GDP, however, Mozambique and Uganda were revealed to be outliers in terms of GDP – due to too low GDP and large losses in GDP due to war – therefore we have no results in terms of GDP. Other outliers in terms of GDP are Zimbabwe (due to a too low GDP), Botswana (due to a too high GDP growth) and Rwanda (due to war). For Kenya, Namibia and South Africa the effects on GDP quite small, while in terms of the birth rate we find no effects. Swaziland on the other hand experienced a drop in the birth rate and no effects on GDP.

Summing up, the effect on GDP seems to be enforced by a non-decline in terms of the birth rate, while countries where the birth rate decreased show smaller or no effects in terms of GDP. The only exception to this finding is Malawi, where we find no effects in terms of birth rate, but quite large effects in terms of GDP.

CHAPTER

4

The Impact of the 1918 Spanish Flu Epidemic on Economic Performance in Sweden ††

4.1 Motivation

Our results from the last two Chapters reveal that the results of HIV on African countries differ from country to country. The reasons for this are twofold. Firstly, there are many channels through which infectious diseases affect the economy. Secondly, HIV is a long-term pandemic which has accompanied us for decades. Therefore, measures were taken to control the disease and to reduce the (economic) consequences. In order to gain a better insight on how infectious diseases affect the economy, we concentrate on one country and turn to a different disease which occurred very rapidly.

In 1918 the world is hit by the Spanish Flu. Estimates suggest that 500 million individuals worldwide were infected by the virus, and that 50-100 million people died in the aftermath of an infection between 1918 and 1920 (Johnson and Mueller 2002). Unlike when customary strains of influenza circulate the world, the majority of the victims of the Spanish Flu were healthy young people in the age interval 15-40 – not frail patients, nor children or elderly.

While much has been written about the medical causes of the Spanish Flu,

††This Chapter is based on Karlsson, Nilsson and Pichler (2012).

the origins of the virus and its connection to more recent pandemics, such as the 2006 bird flu (see e.g. Tumpey et al. 2005, Bos et al. 2011), limited attention has been given to the societal and economic effects of the epidemic. What are the economic consequences following from such a health shock affecting mainly the population of working age within a very short time window?

Using administrative data from Swedish regions, we employ an extension of the standard difference-in-differences (DiD) estimator to exploit the differing mortality rates across Swedish regions. Focusing on Swedish regions has several advantages. First, the variation in flu mortality is high across counties. Almost one percent of the Swedish population died from the Spanish Flu, but there were important regional differences (Åman 1990). For instance some counties experienced more than twice the flu mortality rate of others. We use this variation to examine the impact of the pandemic on earnings, capital returns and poverty.

Second, many key economic indicators are available from Swedish administrative datasets and they are consistently collected across regions and time, allowing for precise estimates.[22] Hence, the data allow us to estimate the effects of the influenza on a number of economic outcomes while carefully checking key methodological assumptions.

Thirdly, Sweden did not take part in the World War I, during which the flu pandemic started. In this way we reduce the risk of confounding effects of the pandemic with disturbances related to the war. Obviously, Sweden was affected by the war in many ways. However, in a non-belligerent country there are no other major shocks to mortality coinciding with the disease. Finally, Sweden is a unitary state and a very homogeneous country and thus there is little need to worry about internal cultural differences or asymmetric responses in regional institutions (cf. Tabellini 2010, Acemoglu et al. 2003).

Our empirical results support the prediction from endogenous growth theory that there will be slower growth in the economy during a transition period after the pandemic. However, in other parts, our empirical results are difficult to reconcile with standard theoretical models. Most importantly, we do not observe the immediate increase in GDP per capita which one would expect as a result of capital deepening. Moreover, there is an apparent redistribution between capital and labor taking place, which suggests that the impact of the pandemic goes beyond what standard growth theory would predict. In the

[22]It is well known in the literature on pandemics that a death caused by influenza was sometimes reported as pneumonia mortality in death records. However, the correlation between influenza and pneumonia mortality at the county level transpires to be quite weak. We interpret this as an indication of the quality of the data and that the detailed instructions sent from national authorities to health personnel on how to verify the cause of death (see e.g. Statistics Sweden 1911) served its purpose and that the correct disease was, in fact, recorded.

discussion of this Chapter, we make an attempt at explaining these findings within the context of a growth model.

This Chapter is structured as follows. Section 3 presents a theoretical model, of how the Spanish Flu affects the economy and reviews previous empirical studies on this topic. In Section 3 we depict the economic environment in Sweden at that time. Section 4 informs about our econometric approach. In Section 5 we present our data. The results appear in Section 6 and are discussed in Section 7.

4.2 The Spanish Flu Pandemic: Facts, Theory and Empirical Evidence

The first official reports on the 1918 flu came from Spain; hence its popular name.[23] Upon reaching the European continent, the spread of the pandemic was accelerated by increased troop movement due to the war (Patterson and Pyle 1991). Among researchers in medical history there is consensus that the disease ran its course in three to four waves. The first wave was in the spring of 1918, with the disease returning in the fall of the same year and again in 1919. The last wave occurred mainly in Scandinavia and some islands in the South Atlantic.

An interesting feature of the second wave of the pandemic is that it took the world by complete surprise. The first wave of the pandemic had such a low mortality rate that experts doubted whether it was influenza at all. For example, in the summer of 1918, Little et al. (1918) conclude

> *we wish to point out that although this epidemic has been called influenza for the want of a better name, yet in our opinion it cannot properly be considered such for the following reasons:*
>
> 1. *The clinical course, though similar to that of influenza, is of very short duration, and there is, so far as we have observed, an absence of relapses, recurrence, or complications [...]*

This is but one example of how medical experts were confused by the *mildness* (!) of the influenza during the first wave, and consequently reluctant to accept it as such. In addition, as the spread of the virus halted in the late summer of 1918, many observers concluded that the epidemic had disappeared (Barry 2005). Contemporary accounts by Swedish doctors also suggest that the first wave was very mild and that there were conflicting views of whether the disease was influenza or a new type of pneumonia (Petrén 1918a,b).

[23]The reason why the first report came from Spain is likely related to the fact that the country did not take part in World War I and at the time had an uncensored media.

This is in stark contrast to the second wave of the Spanish Flu with exceptionally high mortality rates. During a normal influenza epidemic, approximately 0.1 per cent of all infected individuals perish. In comparison to this case fatality rate, the second and most severe wave of the epidemic in the fall of 1918 was 5 to 20 times more deadly. The main reason why the Spanish Flu was so extraordinarily aggressive is that the virus not only attacked the bronchus, but also the lungs, leading to many people dying from pneumonia (Morens and Fauci 2007). The incubation time and the time between infection and death was very short. According to Taubenberger and Morens (2006), most deaths occurred 6-11 days after the outbreak, but there is evidence that some deaths occurred as early as two days after infection (Åman 1990). What furthermore characterizes the disease is the heavy toll among young adults. It is estimated that around half of the death toll was paid by individuals between 15 and 40 (Simonsen et al. 1998). This is unusual and unlike other (influenza) diseases, which typically exhibit a U-shape in the mortality distribution over age groups, the Spanish Flu had a W-shaped distribution over age.

4.2.1 Theoretical Perspectives

From a purely economic point of view, we may think of the Spanish Flu pandemic as labor supply a shock to the economy, which on the other hand leaves physical capital intact. In order to generate hypotheses for how our outcome variables may react to the pandemic, we briefly review the macroeconomic literature on economic growth. Since we cannot observe GDP at a regional level, we look at the returns to capital and labor. Moreover, we are concerned with distributional effects and therefore look at poverty.

Our point of departure is a standard two-sector model (Lucas 1988). Consider an economy with competitive markets. In each location, there is a large number of production units producing a homogeneous final good. Preferences over (per capita) consumption are given by

$$\int_0^\infty U\left(c\left(t\right)\right) e^{-\rho t} \mathrm{d}t \tag{4.1}$$

where ρ is the rate of time preference. Let $h\left(t\right)$ denote the skill level (human capital level) of a typical worker and $u\left(t\right)$ be the fraction of non-leisure time devoted to goods production. Then $1-u\left(t\right)$ is the effort devoted to the accumulation of human capital. It is assumed that the growth of human capital takes a simple form as

$$\dot{h} = \delta\left(1-u\right)h \tag{4.2}$$

where parameter δ is positive.

The output per capita $y(= Y/N)$ depends on the per capita capital stock, $k(= K/N)$, the effective work force uh, and the average level of human capital in the region $\bar{h}$:

$$y = Ak^{\beta}\left[uh\right]^{1-\beta}\bar{h}^{\gamma} \tag{4.3}$$

where parameter β is the income share of physical capital, and parameter γ is positive and captures external effects of human capital. The accumulation of physical capital is assumed to take the form

$$\dot{k} = y - c \tag{4.4}$$

In equilibrium, $h = \bar{h}$ because all production units within the region are treated as being identical. Substituting this into the production function and solving the maximization problem, one gets the social optimum. However, we want to solve for the competitive equilibrium here, which will be done by deriving first order conditions taking the whole path of $\left\{\bar{h}\left(t\right) : t \geq 0\right\}$ as given. Thus, the current-value Hamiltonian may be written as

$$\begin{aligned} H\left(k, h, \theta_1, \theta_2, c, u; A, \sigma, \beta, \gamma, \delta, \left\{N\left(t\right), Y\left(t\right) : t \geq 0\right\}\right) = \frac{1}{1-\sigma}\left[c^{1-\sigma} - 1\right] \\ + \theta_1\left[Ak^{\beta}\left[uh\right]^{1-\beta}\bar{h}^{\gamma} - c\right] + \theta_2\delta\left(1-u\right)h \end{aligned} \tag{4.5}$$

where θ_1 and θ_2 are the co-state variables for k and h respectively. Things taken as given are put after the semicolon in the Hamiltonian.

The first order conditions are thus given as follows:

$$\frac{\partial H}{\partial c} = c^{-\sigma} - \theta_1 = 0 \tag{4.6}$$

$$\frac{\partial H}{\partial u} = \theta_1\left(1-\beta\right)Ak^{\beta}u^{-\beta}h^{1-\beta}\bar{h}^{\gamma} - \theta_2\delta h = 0 \tag{4.7}$$

$$-\dot{\theta}_1 = -\rho\theta_1 + \beta k^{\beta-1}\theta_1 A\left(uh\right)^{1-\beta}\bar{h}^{\gamma} \tag{4.8}$$

$$-\dot{\theta}_2 = -\rho\theta_2 + \left(1-\beta\right)h^{-\beta}\theta_1 Au^{1-\beta}k^{\beta}\bar{h}^{\gamma} + \theta_2\delta\left(1-u\right) \tag{4.9}$$

$$\dot{k} = y - c \tag{4.10}$$

$$\dot{h} = \delta\left(1-u\right)h \tag{4.11}$$

Boucekkine and Ruiz Tamarit (2005) present an analytical solution to this problem for the special case where $\sigma = \beta$. However, before proceeding we derive the factor returns which are crucial in our empirical analysis. Since the economy is competitive, we assume that input factors earn their private

marginal products. Hence, we have

$$r = \frac{\partial Y}{\partial K} = \beta y k^{-1} \tag{4.12}$$

$$w = \frac{\partial Y}{\partial N} = (1-\beta)\, y \tag{4.13}$$

and thus capital returns per capita are given by $rk = \beta y$. A population shock has a direct impact on $k = K/N$ and may also lead to adjustments in consumption c and the allocation of workers to the production of final goods, represented by u. Both these control variables may be adjusted instantaneously, but only u has an instantaneous effect on production and factor returns. Thus, we may write:

$$\frac{\mathrm{d}r}{\mathrm{d}N} = \beta(1-\beta) A K^{\beta-1} (uh)^{1-\beta} N^{-\beta} h^{\gamma} + \beta(1-\beta) A K^{\beta-1} (hN)^{1-\beta} u^{-\beta} h^{\gamma} \frac{\mathrm{d}u}{\mathrm{d}N} \tag{4.14}$$

$$= (1-\beta)\left(\frac{1}{N} + \frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}N}\right) r \tag{4.15}$$

$$\frac{\mathrm{d}w}{\mathrm{d}N} = -\beta(1-\beta) A K^{\beta} (uh)^{1-\beta} N^{-1-\beta} h^{\gamma} + (1-\beta)^2 A K^{\beta} (uN)^{-\beta} h^{1-\beta+\gamma} \frac{\mathrm{d}u}{\mathrm{d}N} \tag{4.16}$$

$$= \beta\left(-\frac{1}{N} + \frac{1-\beta}{\beta}\frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}N}\right) w \tag{4.17}$$

Now consider the responses of labor and capital returns expressed as elasticities:

$$\frac{\mathrm{d}(rk)}{\mathrm{d}N}\frac{N}{rk} = \frac{\mathrm{d}w}{\mathrm{d}N}\frac{N}{w} = \beta\left[-1 + \frac{1-\beta}{\beta}\frac{\mathrm{d}u}{\mathrm{d}N}\frac{N}{u}\right] \tag{4.18}$$

Hence, the immediate impact of the shock is equivalent in the two factor returns, and as long as there is limited accommodation on the part of time spent in education – i.e. $\mathrm{d}u^*/\mathrm{d}N$ is low – the immediate effect of the pandemic is an *increase in earnings* and also an *increase in capital returns*, even though the regional *interest rate* is predicted to fall in response to the population shock.

From (4.7), the optimal allocation of labor is given by

$$u^* = \left(\frac{(1-\beta) A}{\delta}\right)^{1/\beta} \left(\frac{\theta_1}{\theta_2}\right)^{1/\beta} h^{\frac{\gamma}{\beta}-1} k \tag{4.19}$$

Apart from k there are two variables in the equation that can potentially be affected by the labor supply shock: θ_1 and θ_2. If the co-state variables were unaffected, we might expect an increase in the number of hours worked in

production, since capital deepening (the increase in k) has increased labor productivity in this sector. However, this direct effect is likely to be counteracted in a reduction in the shadow cost of capital accumulation – represented by θ_1 – and thus the net effect on workforce allocation may be limited. Indeed, Boucekkine and Ruiz Tamarit (2005) show that for parameter values $\sigma = \beta$, the allocation of labor between production and education is not only constant over time, it is also unaffected by a labor supply shock. In this special case, the immediate elasticity of wages and capital returns with respect to the population shock simply equals β: for each per cent excess mortality, we expect a β percent increase in wages and capital returns.

Turning to the medium-term consequences, the imbalance effects after the population shock are analyzed by Boucekkine and Ruiz Tamarit (2005), again for a particular combination of parameter values. Denoting by ω the ratio between physical and human capital, it can be shown that the growth rate of the economy in the aftermath of the pandemic is going to be proportional to

$$\frac{\bar{\omega}}{\omega} = \frac{\bar{k}/\bar{h}}{k/h} \tag{4.20}$$

where $\bar{k}$ and $\bar{h}$ are the values along the balanced growth path. After the epidemic, ω is above the long-run equilibrium value, and thus the growth rate in the economy is lower than otherwise.

We may thus summarize our theoretical predictions regarding earnings and capital returns as follows:

Hypothesis 1 *If the regional economies behave according to the Lucas-Uzawa model, and the accommodation of labor allocation u^* is incomplete, the influenza pandemic can be expected to lead to*

1. *An immediate increase in earnings w and capital returns rk.*
2. *A slower growth rate (in production, earnings and capital returns) during a transition phase after the pandemic.*

Finally, we also make some predictions concerning poverty. The analysis of poverty is complicated for two reasons. First, the theoretical model does not incorporate any worker heterogeneity and thus it is not useful for making explicit predictions for this variable. Second, the pandemic may have two distinct effects on poverty. First, individuals who were dependent on family members for their living might lose this support in the aftermath of the pandemic. This effect is quite immediate and not directly related to the functioning of the economy – even though Boucekkine and Laffargue (2010) show that an increase in the number of orphans may have important distributional consequences in the long term. According to the 1920 census, on average

each worker in Sweden supported one additional inhabitant (Statistics Sweden 1917). Hence, given the lack of a social gradient in flu mortality, and considering that not all dependents who lost their support became poor, we can think of this number as an upper bound to this effect. Second, changes in wages and capital returns – and their distribution within the population – may have given rise to changes in poverty rates. Given the predictions from *Hypothesis 1* above, we expect these changes to cause an immediate reduction in poverty rates, followed by convergence between more and less affected regions during the transition period. Hence, we may formulate the following proposition:

Hypothesis 2 *The initial effects of the pandemic on poverty will be the net effect of two countervailing forces:*

1. *An increase in poverty due to dependents losing their breadwinners; an effect likely to be smaller than one.*

2. *A decrease in poverty due to rising wages and capital returns.*

In the medium term, both effects are likely to lose importance; i.e., we expect to see a closing of the gap in poverty between heavily and less heavily affected regions.

4.2.2 Empirical Evidence

A growing literature tests the so-called Fetal Origins hypothesis, analyzing the consequences of *in utero* exposure on later health and labor market outcomes, focusing in particular on the effects of the Spanish Flu (cf. Almond and Mazumder 2005, Maccini and Yang 2009, Nelson 2010). These studies suggest long-term damage from prenatal exposure to pandemic influenza and that children of infected mothers are more likely to have health problems and experience lower wages as adults than non-affected children.

In this study, however, we are concerned with short- and medium-term aggregate effects of the pandemic. Up to now there have not been many empirical studies estimating this impact. Besides, existing empirical studies face two serious problems. First, there is a lack of reliable data from the time period. Second, identification is difficult due, *inter alia*, to the fact that the flu occurred during and shortly after the World War I.

Brainerd and Siegler (2003) is one of few papers that consider the effects of the influenza on economic growth. They study changes in real personal incomes between 1919/21 and 1930. Due to data restrictions the analysis only focuses on the medium-term effects and does not distinguish whether the effect was due to recovery or economic growth. In any case, findings suggest significant positive effects: states that where hit harder by the flu experienced

a higher income growth rate from 1919/1921 to 1930. From a theoretical point of view, this result might reflect either capital deepening or be driven by increased investment in human capital and higher population growth after the occurrence of the Spanish Flu.

More recently Garrett (2009) analyses the effects of the pandemic on manufacturing wages. Using the same mortality data as Brainerd and Siegler (2003), but having access to wage growth between 1914 and 1919, the study can compare before and after the pandemic, but is only able to estimate effects in the very short term. The paper concludes that the epidemic appears to have had a positive impact on manufacturing wages. However, it is not always clear to what extent the results are attributable to the World War I.

Focusing on India, Bloom and Mahal (1997a) analyze the effects of the Spanish Flu using data on population changes and acre sown per capita in 13 provinces. India was severely hit by the pandemic, with very high death tolls and the epidemic affected various regions of the country quite differently. Bloom and Mahal (1997a) do not find that any relationship between the magnitude of population decline following from the influenza and the area sown per capita across Indian provinces.

In summary there have been some attempts to estimate the economic effects of the Spanish Flu pandemic in the US and India, but there is still no study which rigorously applies methods typically used to conduct causal inference. The main reason appears to be a lack of reliable data. As shown below, Swedish data appear to offer a significant improvement in this regard.

4.2.3 Drivers of the Influenza

It has been argued that the 1918 influenza pandemic represents a good 'natural experiment' – for estimating short term effects (Brainerd and Siegler 2003) as well as for considering the long-run effects of *in utero* exposure (Almond 2006). The facts that have been forwarded to support this claim are a) the *unexpected onset* of the pandemic in 1918 – which rules out behavioral changes in anticipation of exposure; b) its *short duration*: the majority of deaths occurred within a few months only; c) the *large proportion* of the population infected; and d) the *random nature* of influenza prevalence and influenza mortality.

The assumptions underlying Almond's (2006) analysis have recently been challenged by Brown (2010). The main problem, according to Brown, is that US participation in World War I led to selection issues in childbearing in and around 1919: fathers in the "treatment group" are likely to be older, less educated and less healthy than fathers of surrounding cohorts.

Even though Brown raises valid concerns, it is unclear to what extent they apply to our study. The World War I also led to mobilization and subsequent

demobilization in Sweden, but the disruption caused is of less importance when the short-term impact of the pandemic is concerned. Nevertheless, we now briefly discuss the literature on the determinants of the influenza during the pandemic.

Garrett (2008) analyses the determinants of influenza incidence in the U.S. and finds that even though densely populated areas in general have higher influenza mortality, there is no correlation between 1918 *excess* mortality and population density.[24] We find that this result also holds for Sweden (see Table 4.1).

One particularly relevant study is Mamelund (2006) that considers socio-economic determinants of influenza mortality in the Norwegian capital Oslo (then *Kristiania*). Using register data on influenza mortality, Mamelund estimates the importance of variables such as age, marital status, socio-economic status and quality of housing. Although there are significant class differences in influenza mortality, these appear to be driven more by location than by class itself. Marital status also appears to be insignificant. In a related study, Chowell et al. (2008) consider socio-demographic and geographical patterns in the transmissibility and mortality impact of the epidemic in England. They also fail to find an association between influenza mortality and measures of population density or residential crowding.

4.3 Sweden in the early 20th century

Since it is necessary to consider the particular economic environment which Sweden represented when the influenza pandemic struck in 1918, this Section presents an overview of the general economic and political conditions in Sweden during and shortly after the First World War, and provides an overview of the spread of the influenza epidemic in Sweden.

4.3.1 General economic conditions

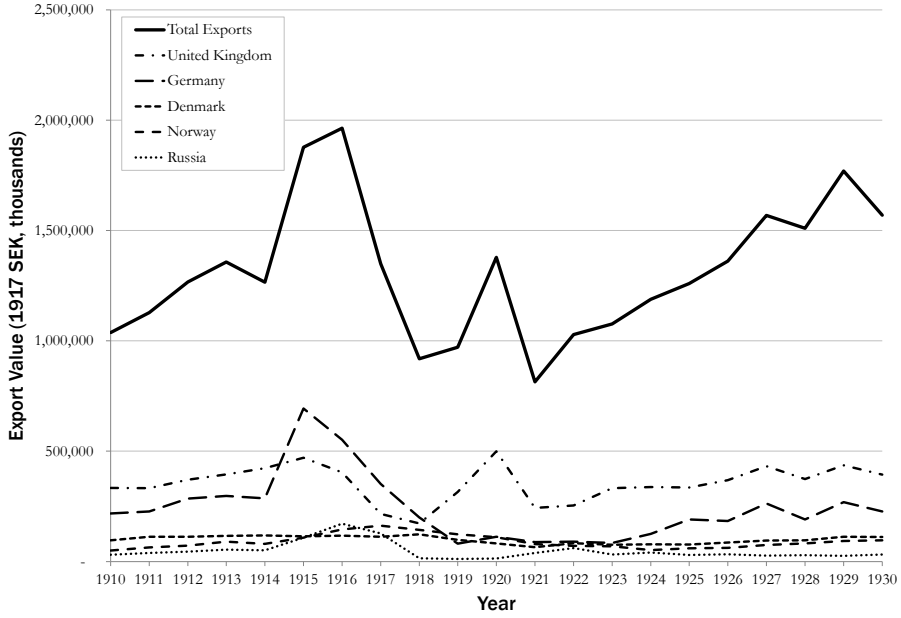
One hundred years ago, Sweden was a radically different society from today. Following a surge in economic liberalizations in the second half of the 19th century, the country had evolved into a modern capitalist state with strong institutions. These reforms included trade liberalization, modern patent laws, and the introduction of joint-stock companies (Bergh 2007). The changes soon gave rise to rapid economic growth.

[24] A related study from New Zealand (McSweeny et al. 2007) concludes that rural areas were less heavily affected by the 1918 influenza; however, the analysis fail to control for the age profile making this finding less informative.

The first half of the 20th century was characterized by rapid industrialization. At the turn of the century, Swedish society was still largely agrarian: according to the 1900 census, 53 per cent of the population earned their living from agriculture and 29 per cent from manufacturing (Statistics Sweden 1907). By 1930, 39.4 per cent of the population still earned their living from agriculture, compared to 35.7 per cent for manufacturing (Statistics Sweden 1936). This structural change occurred at a relatively even pace during these three decades.

Sweden's transformation into a modern industrialized country was largely trade-driven. Figure 4.1 plots Swedish exports to some key trading partners during the 1910-1930 period, expressed in 1917 crowns. Britain and Germany consistently accounted for a large share of Sweden's exports, even if their relative roles shifted back and forth over time. Also, Scandinavian neighbors were important trading partners throughout the period, and their trade with Sweden offered some stability in an otherwise fairly volatile environment. It should, however, be pointed out that the *relative* share of exports in GDP fluctuated much less than the absolute numbers in the figure: exports never went below 14.5 per cent of GDP (1918) or above 21.5 per cent (1913) (Krantz and Schön 2007).

Figure 4.1: Swedish Exports 1910–1930.



In terms of labor market regulations, the period considered falls in between the deregulations that were implemented in the 19th century – such as the abolition of guilds in 1846 and the introduction of free enterprise in 1864 – and the increased regulation that followed the labor movement's rise to power. Thus, wages were relatively flexible and actually dropped in real terms in the 1913-17 period.

4.3.2 Effects of the First World War

At the beginning of the war, Sweden, Norway and Denmark issued identically worded declarations of neutrality. The Swedish army was mobilized shortly after the outbreak of the war, and a ban on exports of arms, ammunition and military equipment was introduced. However, the main disruption to Swedish trade was caused by external forces: the naval blockade imposed by the UK included the entire North Sea. The blockade was very restrictive and, as its implementation was being stepped up, it lead to disruption in Sweden's trade with countries overseas. However, mainly imports were affected (Jörberg and Krantz 1978).

The war also led to increased regulation of the domestic economy. The government was given powers to dispose of resources essential to the military. In 1916, new legislation authorized the government to regulate prices of groceries, fodder, fuel and clothing. This lead to rationing of meat, eggs, butter and fish. However, a black market evolved and thus the regulations were of limited importance in practice (Schön 2010). The period was also characterized by a surge in important social legislation. Even before the war, the first steps had been taken to separate child and elderly care from general poor relief. In 1914, a basic social security system was introduced, including a pension scheme covering the entire population. In the same year, a public committee responsible for unemployment was formed, which was to play an important role in the shaping of active labor market policies (Jörberg and Krantz 1978).

Despite the disruption it brought in some parts of the economy, the war provided a generally favorable economic environment to Sweden. There was a massive surge in exports (iron ore, steel, engineering products) and a huge trade surplus evolved (Magnusson 1996). Shortages in imported fuels led to the electrification of industry production all over the country – improving the competitiveness of Swedish industry. The agricultural sector also benefited from the shortfall in foreign competition. It was only in residential production that investment plunged and remained low throughout the war (Schön 2010).

However, the war gave rise to redistribution between different groups in society. Owners of capital benefited more than workers, and the gains and strains associated with the war were unevenly distributed between different sectors of the economy (Schön 2010). It is important to keep this redistribution in mind, since it was reversed in the post-war slump and thus it might represent a confounding factor with respect to the regional exposure to the Spanish Flu pandemic.

4.3.3 The Roaring Twenties

Having emerged from the World War I largely unscathed, the Swedish economy was subsequently to move on to another decade of rapid economic growth and structural transformation. However, this period of growth was interrupted by a sharp downturn in 1920-21 in which GDP decreased by five per cent in a single year. Interestingly, the industries that had benefited most from the war – such as sawmills and the iron and steel industry – were also the most hard hit by the crisis (Magnusson 1996). There were dramatic increases in unemployment, which reached a level of 25 per cent at the peak of the crisis. However, the recovery was very quick: Swedish GDP increased by 8 per cent in 1922 and the country faced steady economic growth for the rest of the decade (Jörberg and Krantz 1978).

Moreover, the 1920s were characterized by fast growth in real wages: in 1930, they were at roughly twice their 1918 level, and not even in the sharp downturn of 1921 did they stop growing. Thus, the decade was also characterized by a gradual increase in returns to labor compared to capital returns (Schön 2010). Even though unemployment remained relatively high throughout the decade it is believed that the fast growth in wages was partly due to the implementation of the shortening of the working day to eight hours in 1919 (Jörberg and Krantz 1978).[25]

4.3.4 The Spanish Flu Pandemic

With respect to the number of deaths caused, the Spanish Flu is one of the most severe calamities ever to affect Sweden. It killed almost 38,000 individuals, representing almost one per cent of the population. As in other parts of the world, flu prevalence rates were much higher, but generally it is believed that mortality rates amongst those infected approached 2 per cent.

The first case of the Spanish Flu in Sweden was reported in the south in late June 1918. In early August an increasing number of cases are also reported to have died from the flu in the northern provinces. However, as shown by Figure 4.2, until the late summer months of 1918 there was no reason to be concerned about elevated influenza mortality in Sweden. During the first seven months of 1918, 148 influenza deaths were reported, which is below the corresponding

[25]In 1910 the average work week corresponded to 57 hours of work. The working hour act of 1919 stated that a working week should not be longer than six days of work with no more than 48 working hours. Although the act in principle only covered workers in the industry, most workers legally not covered by the new legislation, e.g. employees in the service- and in the public sector, had corresponding working hour restrictions by collective agreements or regulations. For example, government officials had a 45-hour working week in 1920. Following seasonality, farm workers were covered by contracts regulating the maximum number of working hours on a yearly basis (Ryberg-Welander 2000).

figure for 1917 (190 influenza deaths). Yet, once the situation changed in August and September, it did so with a terrifying speed.

Figure 4.2: Monthly Influenza and Pneumonia Deaths. Sweden 1917–20.

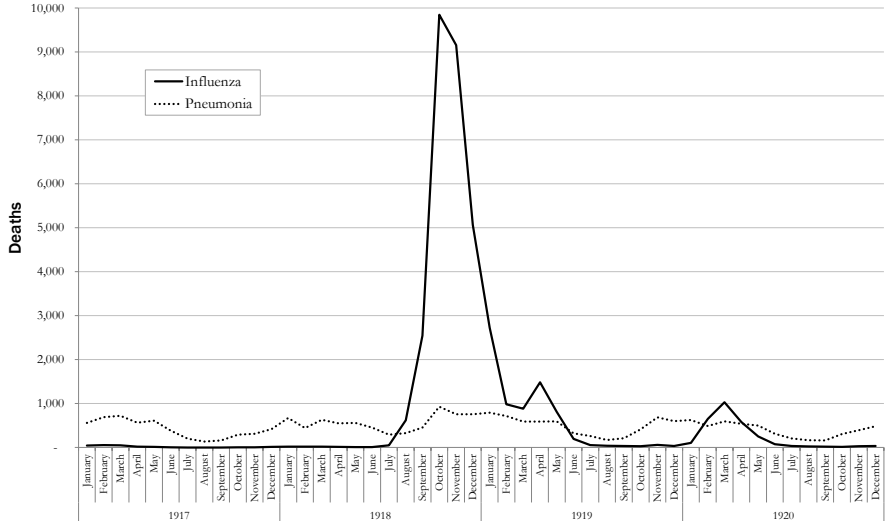


Figure 4.3 shows influenza mortality rates in Swedish counties 1918-1920 (per 100,000 inhabitants). Clearly flu mortality varied widely across counties, with some areas experiencing almost three times higher rates than others. In particular the counties *Jämtland* and *Västernorrland*, were severely hit. The high mortality rates in the remote northern areas have, in part, a demographic explanation as these regions tended to have a young population at the time. However, it has also been hypothesized that the high regional variation in mortality rates may be explained by remoteness, and that people living in these areas had less immunological protection against the virus as they had been less exposed to earlier flu waves. Regarding immunity it has moreover been hypothesized that the W-shaped mortality distribution of the Spanish Flu exhibited in Figure 4.4 may relate to exposure to the Russian flu in 1889-1890.

As discussed above, different industries fared differently during and after the war. Since different regions tend to be specialized in different industries, these fluctuations may become confounding factors. Table 4.1 tabulates all the counties, their influenza exposure and some key statistics from the 1910 census and the year just before the influenza pandemic, namely 1917, when available. Regions are ranked according to their 1918-20 influenza exposure. Interestingly, there is virtually no correlation between sectoral composition and Spanish Flu mortality, suggesting that the spread of the influenza virus was largely unrelated to initial regional economic conditions.

Normal flu waves affecting Sweden typically have their outbreak and peaks in February and March, but the Spanish Flu peaked in October and November. During these two months only, the number of victims of the epidemic reached 20,000 individuals. Another, less severe, wave hit the country in March 1919

Figure 4.3: 1918 Influenza Mortality in Swedish Counties.

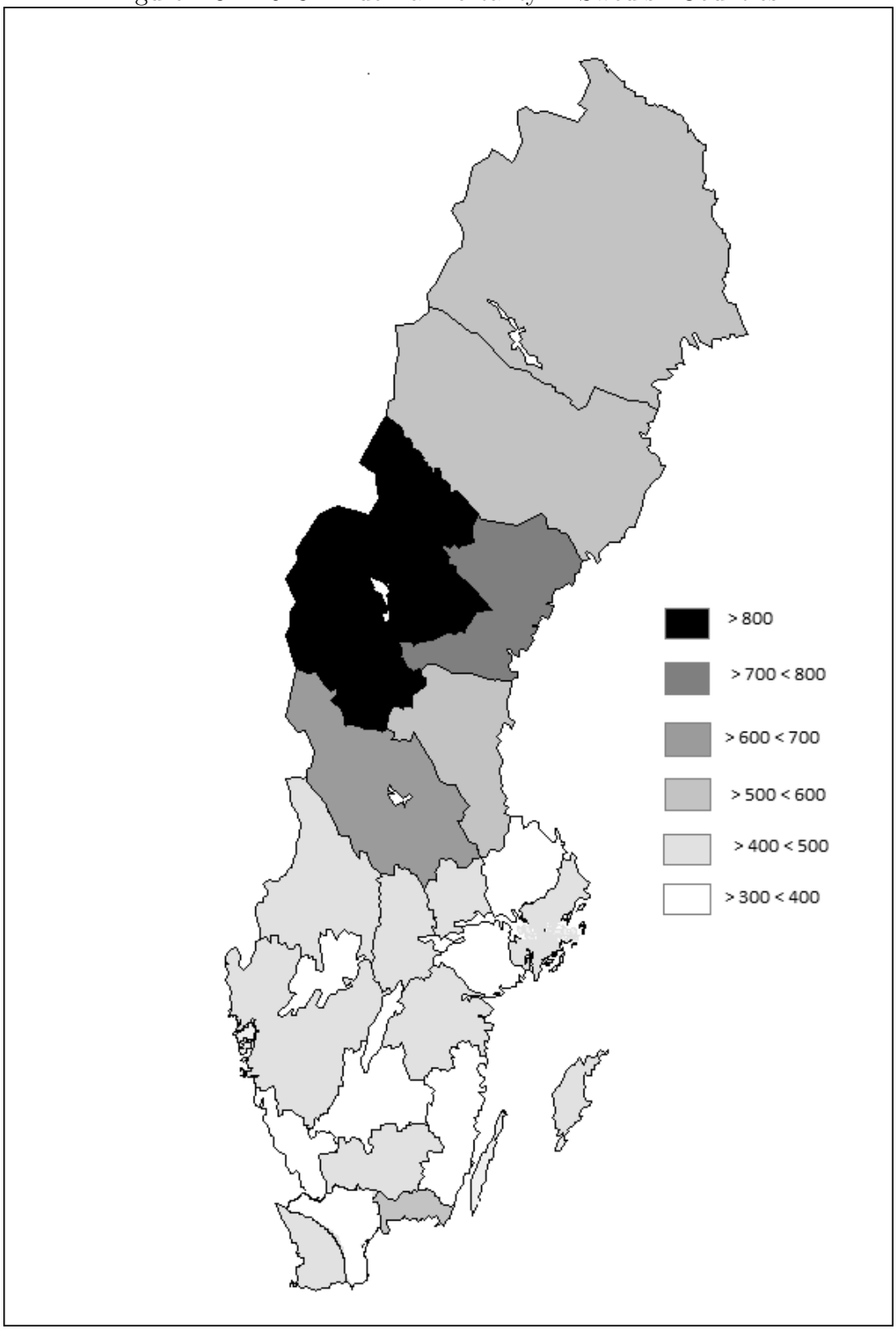


Figure 4.4: Age Distribution of Influenza Mortality.

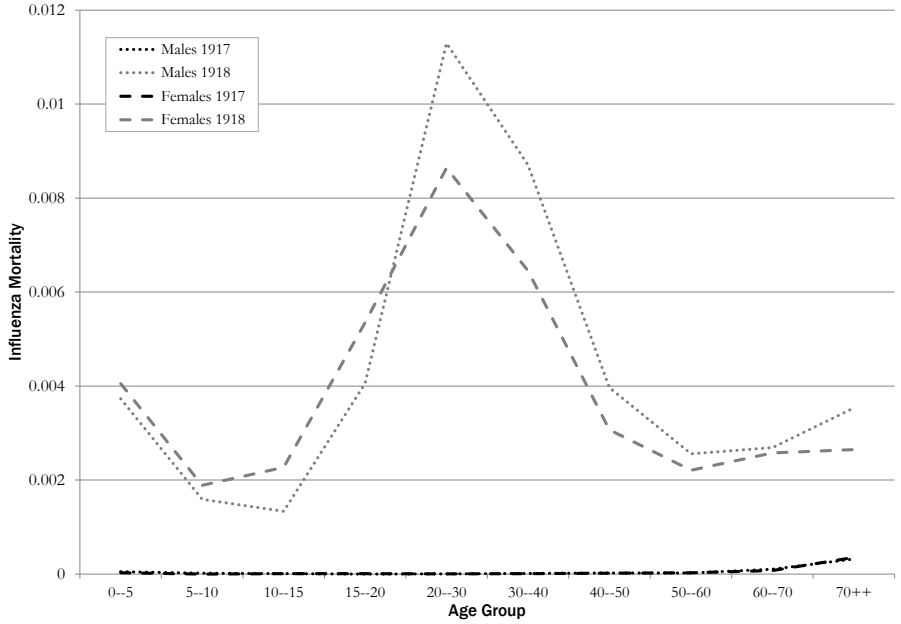


Table 4.1: Treatment Correlations.

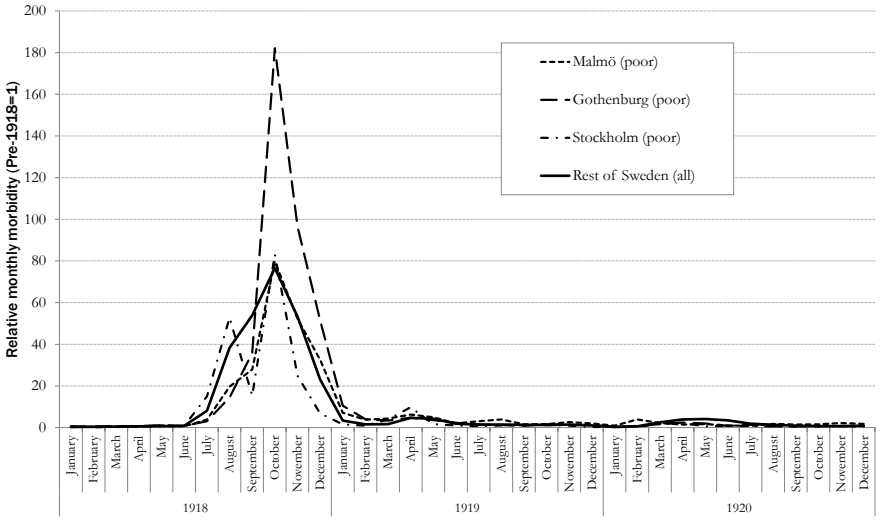
Year	**1918-20**	**1910**				**1917**		
County	Excess Mortality	Population	Agriculture	Manufact.	Commerce	Population	Pop. Density	Earnings per Cap.
Gotlands	445.9	55,217	57.1	16.6	8.0	55,873	17.7	146.4
Södermanlands	463.5	178,568	46.4	27.2	9.4	187,891	27.6	333.1
Kalmar	480.3	228,129	48.1	23.8	9.7	228,998	19.8	183.9
Örebro	494.5	207,021	41.9	34.1	8.6	214,437	23.5	322.0
Jönköpings	504.3	214,454	47.0	28.3	7.3	222,607	19.3	246.7
Uppsala	506.1	128,171	44.9	25.7	7.7	133,506	25.1	301.2
Skaraborgs	510.6	241,284	58.9	18.6	6.5	242,081	28.5	166.1
Kristianstads	545.6	228,307	53.8	21.6	8.4	237,576	36.8	175.2
Östergötlands	552.0	294,179	42.0	29.9	9.8	302,175	27.4	272.5
Älvsborgs	553.3	287,692	53.3	25.0	6.5	297,629	23.4	201.4
Värmlands	562.9	260,135	53.3	25.1	6.8	262,525	13.6	250.0
Kronobergs	567.0	149,654	59.2	19.0	5.8	157,270	15.9	171.2
Västmanlands	577.6	155,920	43.5	31.6	7.8	165,238	24.6	355.8
Hallands	601.4	147,224	52.1	21.4	9.9	147,762	30.0	163.1
Stockholm county	602.0	229,181	35.8	31.8	11.5	230,212	29.7	448.2
Stockholm city	610.7	342,323	0.5	38.1	24.7	413,163	3,642.4	735.2
Göteborgs och Bohus	631.5	381,270	25.2	33.1	18.0	416,843	82.6	406.6
Blekinge	634.7	149,359	40.3	25.8	8.8	148,866	49.4	223.2
Malmöhus	654.6	457,214	30.4	34.3	14.8	481,657	99.7	372.3
Gävleborgs	712.9	253,792	38.1	32.8	11.1	263,989	13.4	328.8
Västerbottens	746.3	161,366	69.3	14.1	4.5	175,031	3.0	171.8
Kopparbergs	748.7	233,873	48.7	31.6	6.7	248,019	8.3	322.4
Västernorrlands	840.3	250,512	47.3	26.9	8.8	262,005	10.3	269.7
Norrbottens	873.7	161,132	54.1	22.4	8.8	177,285	1.7	236.4
Jämtlands	1,017.4	109,851	69.4	12.4	5.2	128,209	2.5	166.8
ρ Flu	1.0	0.01	0.075	-0.086	0.006	0.023	-0.022	0.016

The table shows the standardized excess influenza mortality 1918-20; population size and sectoral shares (public and home sector omitted) according to the 1910 census (Statistics Sweden 1917); as well population, population density (measured in inhabitants per square kilometer) and earnings per capita in 1917. In the last row we present the correlations of the various variables with excess influenza mortality, which are weighted by respective populations in 1910 and 1917.

and new waves appeared until early 1920. Due to the fast spread of the disease in the North, the national government tried to mobilize medical resources to these areas. Moreover, local authorities took actions to limit the spread of the disease and implemented public health measures, such as the banning of public gatherings (Influensakommittén 1924). These actions however had limited effectiveness as the virus was transmitted through the air.[26]

Figure 4.5 provides an overview of the timing of the influenza in Sweden. The curves in the diagram show the ratio between 1918-20 monthly flu incidence and incidence in a 'normal' year. The three dashed curves show the progression of the epidemic among poor people in the three largest cities; and these figures are contrasted with the situation in the entire population in the rest of Sweden. Thus, the figure gives an indication of the socio-economic gradient of the influenza. Accordingly, the poor people in Malmö and Stockholm experienced a slightly lower increase in incidence rates compared to the rest of the country, whereas poor people in Gothenburg were more severely affected.[27]

Figure 4.5: Incidence of Influenza in Different Locations and Socioeconomic Groups.



4.3.5 Assessment

The purpose of this Section has been to give an overview of the environment in which the Spanish Flu pandemic spread in 1918, with a particular focus on potential threats to the identification strategy employed in this Chapter.

[26]There is also detailed documentation on the various treatments that were tested to prevent the spread of the flu in Sweden, see e.g. Influensakommittén (1924).

[27]Amongst poor people in Malmö, the average incidence rate was 9.1 times higher than in a normal year, in Stockholm it was 7.5 times higher, and in Gothenburg 14 times higher. The corresponding figure for the rest of Sweden was 9.7. However, the actual levels of rates are not comparable across locations, since better access to medical services automatically leads to higher recorded incidence and prevalence rates.

We have identified two main threats to the identification strategy which merit special attention. First, even though Sweden soon rebounded from the crisis of the early 1920s, it is clear that the downturn had asymmetric effects between urban and rural areas: in particular, agriculture suffered from a decline in prices when import markets opened after the war. Ironworks and sawmills, which were typically in the countryside, were also particularly badly affected. However, as shown above there was no clear urban-rural divide in the influenza pandemic. A related issue is that the different regions may have been specialized in different sectors of production, and these differences may not be fully captured by the urban/rural dichotomy. Since the industries that benefited most from the war also had a less favorable evolution afterwards, it is essential to establish that the sectoral composition did not lead to counties already diverging during the war. This point calls for a careful investigation of the common time trend assumption for all outcome variables.

4.4 Econometric Approach

4.4.1 Defining the Treatment Indicator

Our analysis is conducted at the level of the 25 Swedish counties. As mentioned, the incidence and mortality of the pandemic exhibit considerable variation across regions. The main assumption underlying our analysis is that the regional exposure to the Spanish influenza represents an exogenous shock and that regions that were affected particularly hard would have followed the same time trend as other regions in the absence of the pandemic. Thus, we define treatment as the total excess regional influenza mortality through the years 1918-20. In our baseline specifications, we furthermore assume that the effects of Spanish Flu mortality is constant over time and a linear function of the excess mortality.

Since the outcome variables are measured annually, we need to correct for the timing of the flu. Most importantly, since the 1918 wave of the epidemic reached its peak only in October and November, it could not have a full effect on the economy in that year. Unfortunately, we do not have monthly mortality data at the county level. However, given that the time period between infection and death was so short (typically 6 – 11 days), we approximate the timing of the fatalities using the timing of influenza incidence.

Thus, we introduce the following notation: yearly flu mortality in county i is denoted m_{it}; and monthly flu morbidity denoted p_{it}^{j} – where t is the year and j is the month. We define both variables as proportions of the county population at the end of year $t-1$. Furthermore, we define the corresponding variables in a 'normal' year (using averages from 1915-17) as m_{i0} and p_{i0}^{j}.

Using these variables, we can define the *effective excess mortality* m_{it}^e in year $t > 1917$ as

$$\begin{aligned} m_{it}^e &= (m_{it} - m_{i0}) \frac{\sum_{j=1}^{12} \left(\frac{p_{it}^j - p_{i0}^j}{2} + \sum_{k=1}^{j-1} \left(p_{it}^k - p_{i0}^k \right) \right)}{12 \sum_{k=1}^{12} \left(p_{it}^k - p_{i0}^k \right)} \\ &= (m_{it} - m_{i0}) \frac{\sum_{j=1}^{12} (12.5 - j) \left(p_{it}^j - p_{i0}^j \right)}{12 \sum_{k=1}^{12} \left(p_{it}^k - p_{i0}^k \right)} \end{aligned} \tag{4.21}$$

In words, $m_{it} - m_{i0}$ is the excess mortality rate on an annual basis. The denominator of the next term normalizes weights such that we adjust for the fact that p_{it}^j captures morbidity and not mortality. In the numerator, we first have the 'excess morbidity rate' of the current month: we divide it by two to correct for the fact that not all cases appear at the beginning of the month. The second term in the numerator captures the cumulative effect of influenza exposure in previous months. For instance, consider any morbidity in year t exceeding the baseline monthly incidence, i.e., $p_{it} > p_{i0}$. If the flu occurs in December the second term reads 1/24, while it would read 23/24 if the flu already hit in January.

Having thus defined the excess flu mortality within a year, we can calculate the cumulative excess mortality at an annual basis. Denoting this variable w_{it}, it is defined as

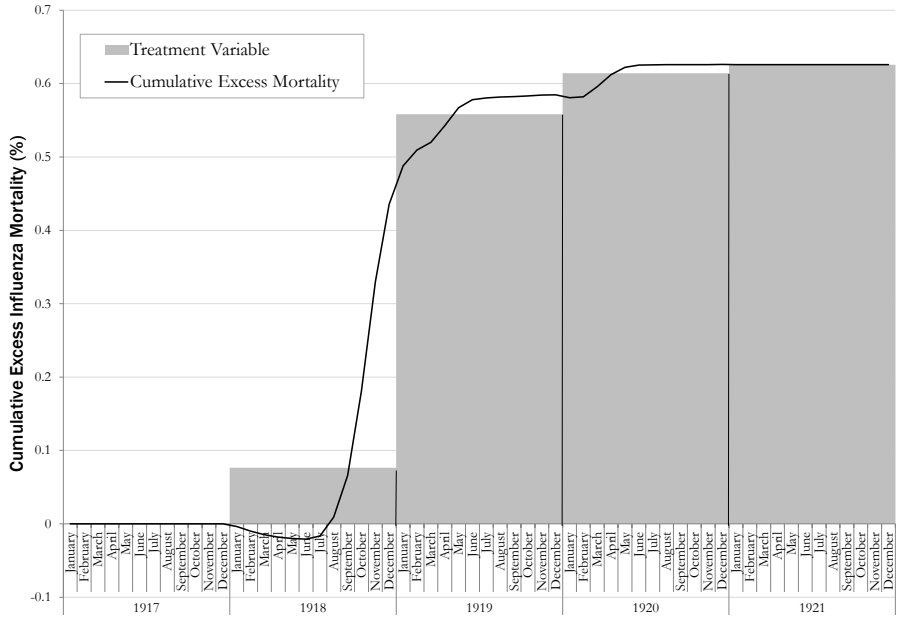
$$w_{it} = \begin{cases} 0 & \text{if } t < 1918 \\ \sum_{j=1918}^{t-1} (m_{ij} - m_{i0}) + m_{it}^e & \text{if } t \in [1918, 1920] \\ \sum_{j=1918}^{1920} (m_{ij} - m_{i0}) & \text{if } t > 1920 \end{cases} \tag{4.22}$$

where, notably, previous years are represented by m_{ij}, not m_{ij}^e: in the following year, we do not need to correct for the timing of period $t-1$ deaths. The treatment indicator is adjusted for the proportion of individuals aged 0-40 according to the 1920 census.

These equations might not seem very intuitive but, in fact, we are simply integrating the number of cumulative deaths over time. This aspect is emphasized in Figure 4.6, where we plot the treatment variable together with the cumulative number of excess deaths. Clearly, the treatment variable adds the integral of new deaths occurring during the current year to deaths that occurred in previous years. This way, the probable impact of these deaths on annual data are captured more convincingly.[28]

[28]We also use alternative treatment indicators, such as an age-adjusted influenza mortality instead of the excess mortality rate, and our result are robust to such alternative specifications. Unfortunately we do not have county data on prime age flu mortality, however, for the city of Malmö we have collected individual level data, which suggests a strong

Figure 4.6: Derivation of the Treatment Variable.



The method we use is an extension of the standard difference-in-differences estimator; our extension is simply that we need to allow for varying treatment intensity (Lechner 2010). Thus, the functional form imposed adds a further assumption to the standard set of assumptions, and it should clearly be formally tested.

4.4.2 Identification

Our empirical analysis crucially rests on the exogeneity assumption, i.e., that the regional exposure to the influenza pandemic was essentially random, and in particular, not correlated with potential outcomes. This assumption is not directly testable, but since it is essential for identification, we have exposed it to a battery of indirect tests, which are described in detail below.

Visual inspection of time trends. The common time trend assumption appears more plausible if one can show that regions with different exposure to the influenza have moved together in the past. Thus, we split the sample into two groups and plot time trends and confidence intervals for all outcome variables considered. If the trends of the two groups diverge already before the 'treatment' in 1918, this is evidence suggesting that the common time trend assumption is not warranted. This graphical test will be performed before the main regressions.

Placebo regressions. By counterfactually assuming that the influenza pandemic hit Sweden between 1915-17 instead of 1918-20, we get an indirect test of not only the common time trend assumption: a placebo regression

correlation between prime age male flu mortality and overall flu mortality.

also tells us something about the statistical properties of the estimator. An insignificant but precisely estimated placebo coefficient suggest that we have acceptable size, whereas an estimate which is significantly different from zero either suggests that the common time trend assumption is violated, or that false positives is an issue. False positives may arise whenever the standard errors are downward biased – for example due to temporal or spatial autocorrelation – and the placebo regressions thus represent a useful test as to whether our dataset suffers from any of these problems. The specification applying a placebo approach is included as one of our regressions.

Relating influenza exposure to pre-influenza covariates. This test goes beyond what is actually necessary for the DID estimator to work, since it is not required that counties are at the same levels before the intervention – only that they follow common time trends. Nevertheless, given the geographical gradient in the influenza, there is the concern that our estimates are confounded by differences in the sectoral composition of the economy and other distinct traits of the pre-influenza regional economies. Thus, we calculate the correlation between our 'treatment' variable and various economic indicators. Doing so, we clearly face a multiple testing problem: random variation in these variables would sooner or later lead to us finding a strong correlation with some covariate. Nevertheless, for identification we require that the correlation between influenza exposure and these additional variables is negligible. Section 3.4 discusses and Table 4.1 presents the results for this test.

Region-specific time trends. In a separate set of specifications, we allow the regions to diverge over time by including region-specific time trends. Including these trends demands more of the data since more parameters need to be estimated and since there is a risk of multicollinearity with our treatment indicator. For this reason, it may be expected that results are somewhat weaker when these trends are included. Hence, we interpret point estimates which do not deviate significantly from our baseline specification as evidence that the estimated effect may indeed be interpreted as causal. The specification applying region-specific time trend is included among our other regressions.

Inclusion of covariates. The DID estimator does not allow inclusion of endogenous variables such as covariates that are possibly also affected by the influenza. However, there are some variables that are plausibly exogenous. It seems reasonable to assume that influenza morbidity (i.e. infections) was exogenous, and also the economic performance of vital trading partners can be assumed to be exogenous from the point of view of the regional economy. If the treatment indicator is truly exogenous, it should be robust to the inclusion of such covariates. However, we take this analysis one step further by also considering covariates which are potentially endogenous. This obviously gives

rise to a 'bad control problem' (Angrist and Pischke 2008), so these estimates need to be interpreted with caution, but it does seem reasonable to check whether results are robust to the inclusion of possible confounders such as birth rates, population density, degree of urbanization, and internal migration. Specifications including additional covariates are included in our regressions.

Collapsing regions. There are several issues related to the spatial structure of the dataset. First, there is the already mentioned problem of spatial dependence of various kinds. Second, we expect migration of workers and capital to even out some of the impact of the pandemic. Third, capital returns are typically registered in the county of the capital holder, which does not have to be the same region as the region where the capital is located. All of these issues can be addressed to some extent by collapsing the counties into larger geographical units. Thus, in a separate set of specifications we collapse the 25 counties into six 'super-regions' with approximately one million inhabitants each. Migration movements between these larger units are much smaller than between the original regions – and thus this alternative specification provides a useful test of whether our results are driven by these other factors.[29] The additional estimations using data for larger regions are presented in Appendix C.

4.4.3 Empirical Specification

For all outcome variables considered the main baseline specification is

$$y_{it} = \alpha_i + \beta w_{it} + \lambda_t + \epsilon_{it} \tag{4.23}$$

where y_{it} is the outcome variable (i.e. capital returns, earnings or poverty), α_i is a county fixed effect, w_{it} is our treatment indicator, λ_t is a year fixed effect, and ϵ_{it} is a residual disturbance. It is straightforward to show that an OLS estimate of β captures the treatment effect if standard assumptions are fulfilled.

It is well known that the DID estimator is sensitive to functional form assumptions. In our case, the natural alternatives are to use either levels or logarithms of the outcome variables. Since the counties are at very different levels at the outset with respect to the outcome variables, a logarithmic specification seems preferable. However, as a robustness check Appendix C also provides estimates for the outcome variables specified in levels.

In an alternative set of specifications, we allow the impact of the influenza pandemic to vary over time:

[29] We also run a specification including a spatial lag of the treatment variable – an indication of the flu mortality in neighbor regions weighted by distance.

$$y_{it} = \alpha_i + \beta w_{it} + \gamma w_{it} \mathbf{1}(t > 1920) + \lambda_t + \epsilon_{it} \tag{4.24}$$

where γ captures treatment effect heterogeneity over time, and $\mathbf{1}(t > \tau)$ is a dummy variable indicating that the year is after 1920.

The placebo regression will take a very similar form:

$$y_{it} = \alpha_i + \delta w_{i,t+3} + \lambda_t + \epsilon_{it} \text{ if } t < 1918 \tag{4.25}$$

In words, we estimate the 'effect' of a counterfactual placebo epidemic, which is assumed to have occurred in the years 1915-17 with the incidence rates of 1918-20. If the placebo parameter δ is precisely estimated and close to zero, it can be seen as evidence for the common time trend. Moreover, it will give us an indication of whether spatial autocorrelation is a problem in the dataset.

4.4.4 Estimating Standard Errors

Inference in DID models has attracted considerable attention in the literature over the past decade. Since our estimates are based on relatively long panels, particular attention needs to be devoted to autocorrelation.[30]

In a seminal paper, Bertrand et al. (2004) discuss the problems associated with autocorrelation in difference-in-differences studies and compare different solutions. One solution which is not discussed in their paper, but outlined by Wooldridge (2009), Stock and Watson (2008) and Arellano (2003), is to use robust standard errors in a fixed effects specification. This combination, which we use in our baseline specifications, is equivalent to clustering at the regional level and thus deals with the autocorrelation problem.

As an additional robustness check, we also reduce the time dimension into five time periods. The estimating equations remain the same as those above, but we now use a collapsed version of the outcome variable, defined as follows:

$$\tilde{y}_{it} = \begin{cases} \frac{1}{T_0} \sum_{s=t_0}^{1917} y_{is} & \text{if } t = 1917 \\ y_{it} & \text{if } t \in [1918, 1920] \\ \frac{1}{T_1} \sum_{s=1921}^{t_1} y_{is} & \text{if } t = 1921 \end{cases} \tag{4.26}$$

where T_0 is the number of time periods before 1918; t_0 is the first year covered by the panel; T_1 is the number of time periods after 1920, and t_1 is the last year covered by the panel. The treatment variable $\tilde{w}_{it}$ is defined analogously:

[30]Since we use data aggregated at the regional level throughout, common group errors as discussed by Donald and Lang (2007) are unlikely to represent a major problem.

$$\tilde{w}_{it} = \begin{cases} 0 & \text{if } t = 1917 \\ w_{it} & \text{if } t \in [1918, 1921] \end{cases} \tag{4.27}$$

Thus, we require estimated effects to be robust to this change in specification.[31]

4.5 Data and Variables

Our analysis of the economic effects of the Spanish Flu is conducted at the level of counties (Swedish: *län*). The data comes from high-quality administrative records. Sweden has a long tradition of collecting official statistics. Statistics Sweden was founded in 1858 and from 1911 onwards the bureau published the series *Sveriges officiella statistik*, divided into nine topics providing information on various issues, on a yearly basis. In addition, most public authorities have a convention of providing official statistics related to their activities.[32]

There are two sources of county-level influenza statistics available for Sweden and they differ to some extent. We use data from Statistics Sweden, which are generally believed to be of high quality and more accurate compared to the influenza statistics provided by *Medicinalstyrelsen* – the authority responsible for national health services at the time. Medicinalstyrelsen's data tend to underestimate the number of cases and also report deaths by place of death and not place of residence (Hyrenius 1914). Statistics Sweden, on the other hand, implemented more detailed and stricter reporting procedures in 1911, generating more complete death cause statistics and improving the reporting from rural areas (Hultkvist 1940).[33] With respect to accuracy, reporting from urban areas were most likely, however, superior to reporting from rural areas, although it should be noted that special reporting procedures applied to deaths

[31] A third alternative would be to rely on the GLS estimator originally suggested by Kiefer (1980). In a recent paper, Hausman and Kuersteiner (2008) analyze the properties of this GLS estimator. Their main conclusions are that a FGLS procedure generally outperforms procedures where the time dimension is reduced by aggregating observations. Even though their size correction is promising, we decided not to follow that route here. The estimated correlation matrix exhibited positive autocorrelation in the short term but negative autocorrelation in the long term – and, thus, standard errors often turned out smaller than in the original OLS specification.

[32] Official data for the time period covered in our analysis is available in hard copies and sometimes as scanned documents. The information used in this Chapter has been digitalized by the authors and their research assistants.

[33] Before 1911 there was no clear guidance on what could be defined as a death cause and how to record the main cause of death and often several death causes were reported in turn reducing data accuracy (Hyrenius 1914). The new procedures likely also improved preciseness and the correctness of death cause statistics as the main death cause of a deceased hereinafter always was decided upon by a doctor. Clergymen had to make monthly reports on the likely cause of death of persons in cases where no doctor had been involved. These notes were then reviewed and confirmed by a GP who reported the final cause of death to the bureau. For details see the introductory Chapter in Dödsorsaker 1911 (Statistics Sweden 1915).

related to epidemics in both rural and urban areas (Hyrenius 1914).[34]

We use data from Statistics Sweden on county-level influenza deaths reported on a yearly basis. As described below, we use this information together with monthly influenza incidence statistics from Medicinalstyrelsen, to derive our treatment variable. Incidence data are of a lower quality than the mortality data due to the fact that the patient had to visit a physician to be recorded. However, doctors were obliged to report verified cases of the flu (Influensakommittén 1924) and governmental historical records (see e.g. Influensabyrån 1919) suggest that people did visit health care centers when they had the flu and that the pandemic clearly increased the demand for GPs.

In baseline regressions we use yearly data for the time period 1912-1930 and focus on three economic outcomes. The first outcome variable is *capital incomes* per capita defined as incomes from e.g. asset yields, rents and dividends taken from official tax records (Statistisk Årsbok).[35] We also use *earnings* per capita, referring to all taxed earnings from employment and pensions per capita collected from the same source.[36] From 1903 it was mandatory for all adults in Sweden to declare their incomes to the tax authorities. Everyone had to state their yearly earnings (including payment in kind and pensions), after deductions of pension contributions and for business expenses, and capital incomes to local tax boards that examined and controlled the declarations and those with an annual income of more than 600 crowns were taxed. Clearly there might be differences between the taxed income amount and actual incomes. However, as discussed by Roine and Waldenström (2009) the administrative routine in Sweden has been very thorough through the twentieth century and Swedish tax data are quite reliable. Moreover, contemporary sources report that the main difficulty was to get accurate information for property taxation rather than incorrectly reported incomes (Statistics Sweden 1921).[37]

The third outcome variable is *poverty rates*, referring to the number of inhabitants in public poorhouses as a proportion of the total population in each region, collected from the yearly publication *Fattigvården*. Following information from Statistics Sweden (1911), people who were not able to support themselves or could not be supported by their family were eligible for the pub-

[34]Special reporting procedures also applied to violent deaths and suicide.

[35]All monetary outcome variables are adjusted to real measures using 1917 as base year. The measure used for adjusting the variables is regional cost of living numbers provided by Statistics Sweden. All results are robust to using the Swedish national CPI, also available from Statistics Sweden.

[36]National pensions have basically always been regarded as taxable income in Sweden. As discussed by Elmér (1960) the amounts were small during the first decades and likely often not even declared.

[37]The standardized self declaration form had to be signed on word of honor. The punishment for submitting incorrect or improper information, which thereby led to that earnings were not taxed, was a fine between four to ten times the amount not reported (Riksskatteverket 2003).

lic poorhouses governed by the municipality.[38] All applicants to poorhouses were carefully registered and exposed to a means test. An individual that was accepted to a poorhouse received housing, clothing, food, medical care and medicine, and the coverage of funeral costs, but were also declared legally incompetent (Rauhut 2002).[39] Statistics Sweden provides information on the number of poor since 1871 when a new law demanding all municipalities to yearly provide statistical accounts is implemented. Information on the total number of poor, but also their sex, age and marital status was systematically reported to the authority using standardized forms (for an example, see Jorner 2008). In order to avoid spurious effects of the influenza working through the denominators of the per capita variables, we use the average county population over the year throughout.

According to the yearly documentation and summary reports from our data sources, all variables seem to have been consistently collected across the time period of interest. Notably, Statistics Sweden implement quality improving changes in their data collection routines from 1910.[40] Importantly, as discussed by Jorner (2008), Statistics Sweden's death causes are classified according to the 1911 nomenclature until 1930 and we have not noted any changes in any of the definitions of the other above indicators that could influence our results.[41]

Table 4.2 provides descriptive statistics for all variables. For the sake of comparability, all monetary variables have been expressed in 1917 crowns according to the average cost of living in the region. We provide averages of the variables for the period *before* (up until 1917), *during* (1918-20) and *after* (1920-30) the pandemic. Concerning the earnings variable – which together with the capital returns variable is based on official tax records – the numbers seem to be well in line with those available from other sources. As can be seen in the table, average taxable earnings per capita were 381 crowns during the 1918–20 period. The corresponding figure for 1920 only is 432 crowns (not shown). According to the 1920 census, male industrial workers earned 1,600 crowns per year on average (females 1,000), whereas agricultural workers typically earned less than 1,000 crowns and female workers in agriculture as little

[38] As discussed by Elmér (1960) the 1913 pension reform reduced the number of old people in the poorhouses significantly wherefore demographic population structure should not explain regional variation in poverty.

[39] The inhabitants of poorhouses could consequently not vote in elections or referendums, nor get married or move away from the municipality.

[40] From 1910 data is more often collected directly from relevant informants rather than from administrative sources. The new routines follows from the investigation presented by the 1905 statistical committee on how Statistics Sweden should accomplish their mission (Jorner 2008).

[41] As described by Elmér (1960) the Swedish pension system was basically unchanged during the period 1913-1936. Recent research also indicates that there have not been any systematic changes in the level of tax avoidance and evasion during the studied time period (Roine and Waldenström 2009)

as 544 crowns.[42] Thus, if one considers the fact that only 49.9 per cent of the population was working, the average earnings of 432 crowns for 1920 seem to be of a similar order of magnitude.

Table 4.2: Summary statistics.

Variable	N	Mean	St. Dev.	Before	During	After
Capital Income (SEK/capita)	475	39.659	42.349	22.593	42.843	49.622
Earnings (SEK/capita)	475	404.124	235.660	276.929	381.107	499.151
Poverty (%)	475	4.615	1.408	4.183	4.043	5.157
Trade Demand (SEK)	475	140.922	44.101	129.786	116.579	159.164
Population (, 000)	475	236.476	98.254	227.957	235.197	242.723
Flu Incidence, Per Capita	475	0.006	0.020	0.000	0.026	0.000
Cum. Flu Incidence, Per Capita	475	0.066	0.056	0.000	0.080	0.105
Cum. Flu Mortality (w_{it}; %)	475	0.395	0.309	0.000	0.469	0.626
Population Density (Inhabitants per km^2)	475	176.360	736.628	162.173	172.428	187.570
Rural Population (%)	475	75.665	19.551	77.496	75.790	74.389
Birth Rate (‰)	475	19.867	3.872	22.391	21.542	17.440
Internal Migration (‰)	475	-0.281	8.172	-0.251	0.852	-0.805
Immigration (‰)	475	1.051	0.669	1.185	1.231	0.881
Emigration (‰)	475	1.972	1.461	2.174	1.226	2.169

The table shows descriptive statistics for the variables, and shows means of all variables before (t<1918), during (t $\geq$ 1918 & t $\leq$ 1920) and after (t>1920) the Spanish Flu pandemic. Incidence (infections) and mortality have been calculated as excess rates.

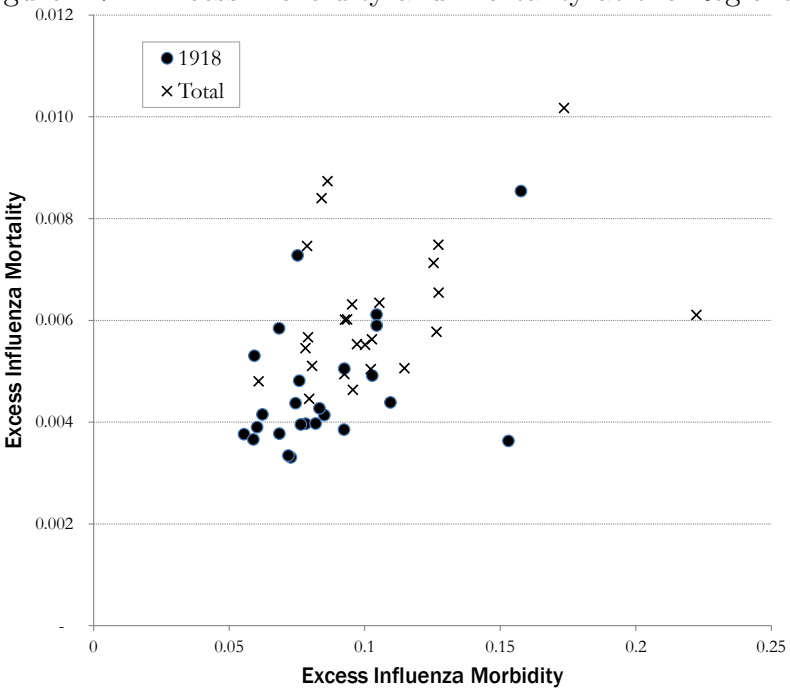
A concern is that the estimated effect of influenza mortality may actually be capturing long-lasting effects of influenza *prevalence*. The growing literature on effects of *in utero* exposure provides but one example of how the effects of the influenza might manifest themselves at the regional level (see e.g. Almond 2006). Thus, we include influenza incidence in some separate specifications as a robustness check. Figure 4.7 shows the relationship between excess morbidity and excess mortality at the regional level. Even though the variables are clearly positively related, they are not as strongly correlated as one might expect: in the year 1918, the correlation coefficient for flu is 0.43.

Another concern regarding the internal validity is the volatility of the world economy during the time period studied. Section 4.3 suggests that the Swedish economy appears to have weathered crises in the surrounding world relatively well. Nevertheless, the Spanish Flu pandemic was preceded by the First World War and the Russian Revolution, and largely coincided with the 1918-19 revolution in Germany (one of Sweden's main trading partners) and the civil war in Russia (including Finland, bordering Sweden). If these and other external events caused disruption to the economy, and if these influences were spatially heterogeneous in a way that coincides with the exposure to the epidemic, then our estimates of the effect of the epidemic may be biased.

In order to check the robustness of our findings we take the volatility of the economic environment into account by also including information on GDP in other countries in some specifications. Information on GDP and population

[42] Own calculations based on Statistics Sweden (1926) and a CPI deflator of 1.6524.

Figure 4.7: Excess Morbidity and Mortality at the Regional Level.



size is available for the 27 countries which together represent virtually all of the Swedish exports of the time. Our trade variable is derived in two steps. First, we estimate a partial gravity function[43], where Swedish exports to other countries were explained with reference to their distance, their GDP and their GDP per capita:

$$\ln(PX_{st}) = \delta_0 + \delta_1 \ln(GDP_{st}) + \delta_2 \ln(GDP_{st}/Pop_{st}) + \delta_3 D_s + \upsilon_{st} \quad (4.28)$$

where PX_{st} are Swedish exports to country s in year t, GDP_{st} is the gross domestic product of country s in year t, Pop_{st} is the population size, and D_s is the distance from Stockholm to the capital of country s. We estimate equation (4.28) using the random effects estimator. The results indicate that distance and total GDP are strongly significant, whereas GDP per capita is marginally significant.

In the next step we generate the variable $\widehat{PX}_{it}$ for each county i. This variable refers to the total exports that would be expected in year t if Sweden were located at the centroid of county i:

$$\widehat{PX}_{it} = \sum_{s=1}^{27} e^{\hat{\delta}_0} GDP_{st}^{\hat{\delta}_1} \left(\frac{GDP_{st}}{Pop_{st}}\right)^{\hat{\delta}_2} D_{si}^{\hat{\delta}_3} \quad (4.29)$$

where D_{si} now represents the distance between county i and country s. Clearly,

[43] Anderson (1979) provides the first theoretical foundation of a gravity trade model; cf. Rose (2000) for an overview of the literature.

$\widehat{PX}_{it}$ has no obvious interpretation in economic terms, partly because equation (4.28) is only half a gravity equation. Nevertheless, we believe that this variable goes a long way towards controlling for asymmetric shocks related to the business cycle and major events in neighboring countries.

4.6 Results

4.6.1 Common Time Trend: Visual Evidence

Our case differs from the standard DID setting in the sense that we have more than two degrees of treatment intensity, and hence counties included in the analysis do not form two distinct groups. However, in terms of the total excess influenza mortality experienced over the entire 1918-20 period, we may distinguish two different strata of exposure. Most counties fall within the range of 440-700 additional deaths per 100,000 population. Above that, there is a smaller group of six counties which experienced between 700 and 1,017 additional deaths. To provide some visual evidence concerning the common time trend assumption we contrast these two groups.[44] In Figures 4.8 to 4.10, counties are weighted by their 1917 population size[45], and all monetary variables are expressed in 1917 crowns (adjusted according to the regional price level obtained from average regional cost of living). The solid curve in Figure 8 pictures growth in capital incomes for counties which were hit particularly hard by the epidemic. The dotted curve plots the corresponding series for the less severely affected counties, while the gray curves show 95 % confidence intervals.

The graphs indicate that the common time trend is a reasonable assumption before the pandemic hit: the curves are quite close and their confidence intervals overlap. However, during and after the pandemic, the two groups diverge for most outcome variables.

Figure 4.9 suggests that more strongly affected counties experienced slower earnings growth, and Figure 4.10 suggests that poverty increased by the pandemic. Both observations might be driven by a change in poverty rates. Thus, we define an alternative earnings variable, where total annual earnings at the county level are divided by the number of inhabitants who are not poor. Figure 4.11 provides visual evidence for the modified earnings variable. The common time trend assumption now appears to be even more plausible than for the original variable, but otherwise no important changes are discernible.

In conclusion, there are no blatant violations of the common time trend in

[44] Allowing for more groups does not change the results, but make the figures more difficult to read.

[45] The results not using any weights are very similar to the results presented in the text.

Figure 4.8: Common Time Trend for Capital Income.

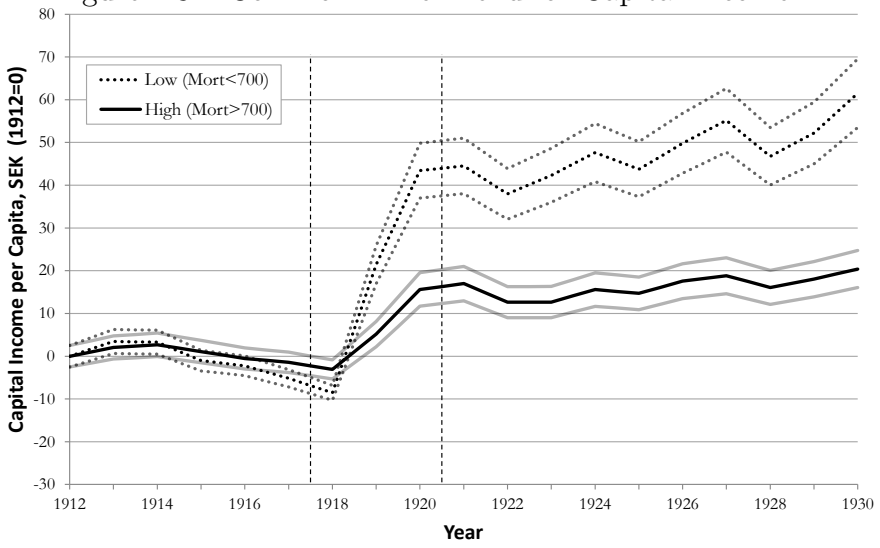


Figure 4.9: Common Time Trend for Earnings.

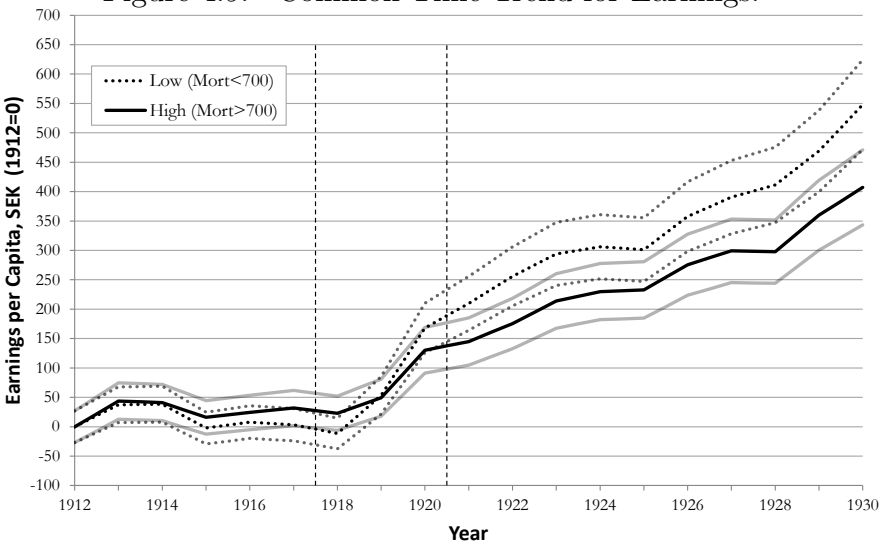


Figure 4.10: Common Time Trend for Poverty.

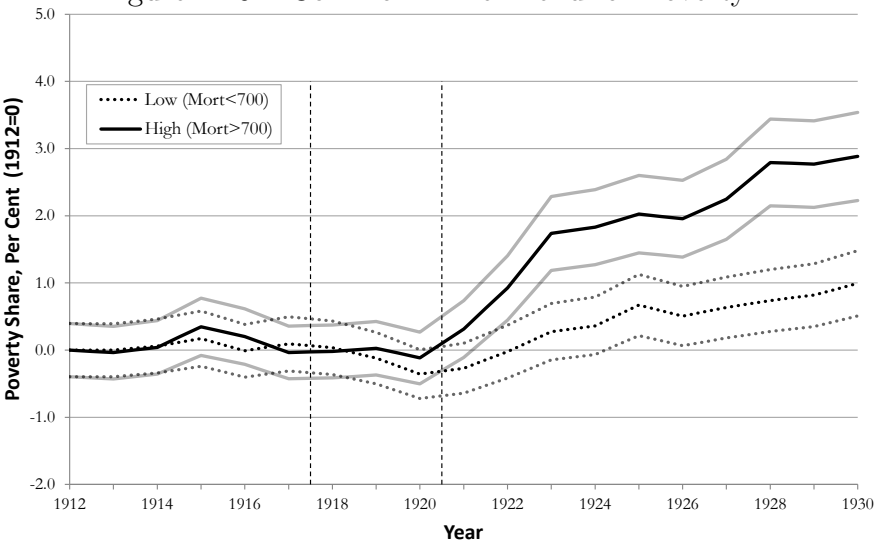
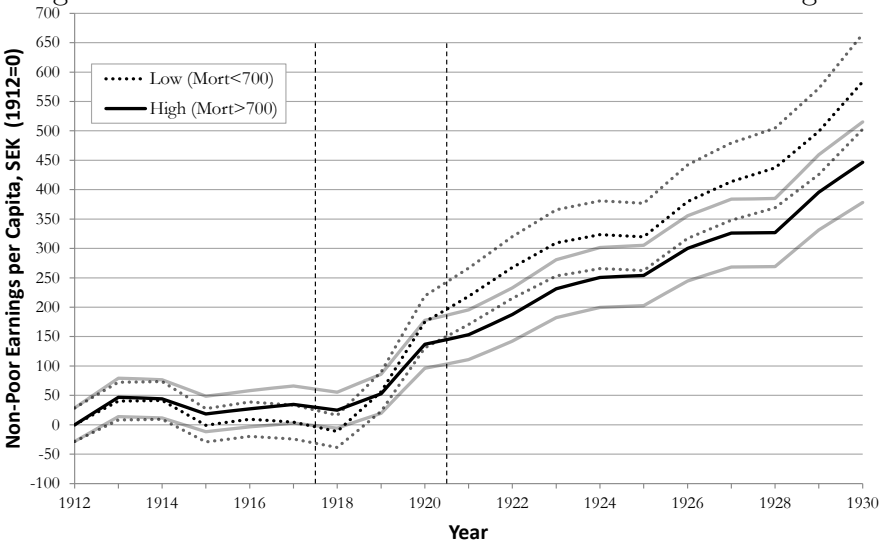


Figure 4.11: Common Time Trend for Non-Poor Earnings.



our data, and the pandemic appears to have had an impact on our outcome variables of interest. Clearly, however, the above evidence is too crude and summaric to provide a reliable estimate of the effects. Hence, we now turn to more rigorous regression-based evidence.

4.6.2 Regression Analysis

Table 4.3, Panel A presents the results for *capital income*.[46] The first column presents the overall effect of the pandemic. According to our estimate, each additional death per 100,000 inhabitants was associated with a reduction in capital income per capita by 0.083 per cent. To get an idea about the magnitude, one may compare the 25th and the 75th percentile, with an incidence of 0.291 and 0.616 respectively. The difference between these two counties would correspond to a reduction in capital incomes per capita by 27 per cent.

In the second column, we contrast the effects *during* (1918–20) and *after* (1921–30) the pandemic (as defined in equation 4.24). Parts of the effect are discernible during the pandemic itself, whereas there is an additional effect kicking in afterwards. In column three we include *regional time trends*. As this variable exhibits a strong correlation with our treatment variable it is not surprising that the estimated effect weakens somewhat. Columns four and five control for cumulative morbidity and current morbidity respectively. The sixth column presents estimates controlling for export shocks (see the discussion before equation 4.28 for further details). Moreover, in column seven we include additional control variables (birth rates, internal and external migration, population density, percent of rural population).[47] Throughout these different specifications the variation in the estimated treatment effect is very limited. Finally, column eight allows for a 'placebo epidemic' (from equation 4.25). This estimate is nowhere near statistical significance and it is very precisely estimated. Thus, our observation from Figure 4.8 is confirmed and the common time trend assumption is maintained.

In columns nine and ten, we collapse the time period into five periods in order to reduce problems related to autocorrelation (see equations 4.26 and 4.27 for details). The estimates clearly indicate that autocorrelation is an issue. Nevertheless, the effects observed after the epidemic are still significant at the five per cent level. Thus, we may conclude that we have found very strong and robust evidence of a substantial immediate effect of the pandemic on capital returns, and fairly strong evidence of a further reduction in capital

[46] In the baseline analysis we take the natural logarithm of all dependent variables. In Appendix C we provide estimates for the outcome variables in levels (Table C-1).

[47] We also perform a regression with migration as a dependent variable to see whether the Spanish Flu affected migration behavior. The treatment indicator turns out to be insignificant suggesting that people were not "fleeing" from the flu.

returns after the pandemic.

Table 4.3: Regression Results, Logarithm of Outcome Variables.

Panel A: Capital income	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	-0.831***	-0.560***	-0.363*	-0.519**	-0.561***	-0.620***	-0.560***		-0.579***	-0.478**
	0.197	0.182	0.202	0.211	0.182	0.167	0.182		0.187	0.189
$w_{it} \times \mathbf{1}(t > 1920)$		-0.317*	-0.223	-0.318*	-0.322*	-0.277	-0.225			-0.266
		0.174	0.193	0.175	0.175	0.169	0.191			0.202
Placebo (w_{it+3})								0.00170		
								0.105		
Cum. Incidence				-0.399						
				0.619						
Incidence					-0.553					
					0.736					
Trade Demand						2.502***				
						0.881				
Constant	3.034***	3.034***	3.034***	3.034***	3.034***	-9.213**	3.321***	3.034***	3.080***	3.080***
	0.0306	0.0306	0.0201	0.0291	0.0305	4.305	0.592	0.0202	0.0292	0.0293
Further Controls	No	No	No	No	No	No	Yes	No	No	No
Regional Time Trends	No	No	Yes	No	No	No	No	No	No	No
Observations	475	475	475	475	475	475	475	150	125	125
R^2	0.958	0.958	0.973	0.958	0.958	0.960	0.961	0.765	0.954	0.954

Panel B: Earnings	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	-0.204	-0.220	-0.186	-0.119	-0.222	-0.226	-0.244		-0.206	-0.215
	0.239	0.202	0.168	0.217	0.201	0.204	0.174		0.200	0.184
$w_{it} \times \mathbf{1}(t > 1920)$		0.0187	0.0342	0.0143	0.0123	0.0221	0.166**			0.0248
		0.0787	0.0990	0.0792	0.0812	0.0779	0.0651			0.0843
Placebo (w_{it+3})								-0.0423		
								0.115		
Cum. Incidence				-0.984**						
				0.364						
Incidence					-0.678					
					0.575					
Trade Demand						0.214				
						0.671				
Constant	5.570***	5.570***	5.570***	5.570***	5.570***	4.523	4.775***	5.570***	5.660***	5.660***
	0.0308	0.0308	0.0156	0.0264	0.0305	3.275	0.285	0.0221	0.0168	0.0169
Further Controls	No	No	No	No	No	No	Yes	No	No	No
Regional Time Trends	No	No	Yes	No	No	No	No	No	No	No
Observations	475	475	475	475	475	475	475	150	125	125
R^2	0.969	0.969	0.987	0.971	0.969	0.969	0.976	0.613	0.964	0.964

Panel C: Non-Poor Earnings	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	-0.166	-0.219	-0.204	-0.0991	-0.220	-0.223	-0.240		-0.188	-0.214
	0.235	0.200	0.168	0.219	0.198	0.201	0.173		0.197	0.182
$w_{it} \times \mathbf{1}(t > 1920)$		0.0614	0.0674	0.0562	0.0560	0.0643	0.199***			0.0668
		0.0730	0.0914	0.0737	0.0750	0.0725	0.0644			0.0783
Placebo (w_{it+3})								-0.0353		
								0.112		
Cum. Incidence				-1.159***						
				0.333						
Incidence					-0.564					
					0.525					
Trade Demand						0.189				
						0.683				
Constant	5.611***	5.611***	5.611***	5.611***	5.611***	4.686	4.868***	5.611***	5.705***	5.705***
	0.0312	0.0312	0.0143	0.0259	0.0310	3.332	0.277	0.0217	0.0171	0.0172
Further Controls	No	No	No	No	No	No	Yes	No	No	No
Regional Time Trends	No	No	Yes	No	No	No	No	No	No	No
Observations	475	475	475	475	475	475	475	150	125	125
R^2	0.971	0.971	0.988	0.974	0.971	0.971	0.978	0.627	0.965	0.966

Panel D: Poverty Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	0.642**	0.0201	-0.174**	0.271	0.0227	0.0399	0.0699		0.312*	0.0452
	0.242	0.0985	0.0813	0.182	0.0940	0.101	0.0988		0.168	0.102
$w_{it} \times \mathbf{1}(t > 1920)$		0.727***	0.635**	0.716***	0.741***	0.714***	0.544***			0.703**
		0.242	0.260	0.245	0.246	0.243	0.187			0.263
Placebo (w_{it+3})								0.0321		
								0.0944		
Cum. Incidence				-2.431**						
				1.025						
Incidence					1.448*					
					0.813					
Trade Demand						-0.832				
						0.630				
Constant	1.440***	1.440***	1.440***	1.440***	1.440***	5.512*	2.086***	1.440***	1.452***	1.452***
	0.0272	0.0272	0.0259	0.0222	0.0277	3.064	0.366	0.0116	0.0197	0.0197
Further Controls	No	No	No	No	No	No	Yes	No	No	No
Regional Time Trends	No	No	Yes	No	No	No	No	No	No	No
Observations	475	475	475	475	475	475	475	150	125	125
R^2	0.554	0.575	0.804	0.648	0.579	0.577	0.759	0.141	0.535	0.593

The table shows four panels with results from fixed effects regressions. In the first specification we regress the natural logarithm of the dependent variable (which varies with each panel) on our treatment variable w_{it}. The second column additionally interacts the treatment variable with a dummy which equals one for data points after the Spanish Flu. The third column includes region-specific time trends. Column 4 (5) controls for (cumulative) flu infection rates. Column 6 includes trade demand as an additional control. For the derivation of this variable, see Section 4.5. Specification 7 includes further controls, including birth rates, migration within Sweden and abroad, population density and percent of population in rural areas. Column 8 presents the placebo regression which is estimated using years before 1918 and the third lead of the treatment variable ($w_{i,t+3}$). The final specifications 9 and 10 collapse the data in order to control for autocorrelation. All regressions include year dummies which are not displayed and are weighted by the population in 1917. The second row presents robust standard errors. As discussed by Wooldridge (2009), Stock and Watson (2008) and Arellano (2003), in a fixed effects model, robust standard errors are equivalent to regionally clustered standard errors. The asterisks represent significance at the following p values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.3, Panel B provides estimates for the *earnings* variable. For this outcome there is much less evidence of a flu effect. The point estimate of the overall effect is -0.2, which, according to our previous comparison, would imply a relative decline of 6.5 per cent in the 75th percentile county compared with the 25th percentile. Importantly, the placebo estimate is smaller and estimated with a similar degree of precision. Hence, the common time trend assumption cannot be rejected, and we may thus conclude that the epidemic appears to have had no effect at all on earnings per capita.[48] Next, Table 4.3, Panel C, presents the results when focusing on earnings of the non-poor population. Normalizing earnings using the non-poor instead of the total population does not change our conclusions: the estimated effect is still insignificant and very similar to our previous estimates.

Table 4.3, Panel D reports results for *poverty rates*. The pandemic appears to have had a strong and lasting positive effect on poverty. The overall effect is estimated at 0.64. Comparing the 25th and the 75th percentile, the difference in flu mortality would give rise to an increase in poverty by 20 per cent. Again, the influenza effect is quite substantial – but it only appears after the pandemic receded. It is important to remember that there is a direct mechanism at work, which has little to do with the functioning of the economy, to the extent that deceased individuals leave dependents behind, who are unable to support themselves. Two pieces of evidence however suggest that this factor is not the main driver of the positive relationship between the influenza and poverty. First, we performed an additional analysis on disaggregated poverty statistics. This analysis shows that the baseline effect is neither driven by widows nor

[48] We also obtained data on agricultural wages. Data are not available for all counties and hence we do not include the results; however, the empirical evidence for this variable suggests that no flu-effect was observed on agricultural wages either.

by orphans. Secondly, a close inspection of Table C-1 Panel C in Appendix C further reveals that dependents cannot be responsible for the entire effect: according to our estimates from the specification in levels, each death caused by the epidemic led to *four* additional poorhouse residents – and, considering the age pyramid in those days, it is implausible that all newly poor would were dependents of a deceased person.

Also for poverty the placebo estimate is insignificant, small and, in relative terms, precisely estimated. Thus, in line with the visual evidence the common time trend assumption seems to be confirmed also in this case. We conclude that the pandemic appears to lead to a large increase in poverty rates in the medium term.

Appendix C provides estimates for the outcome variables in levels (Table C-1). Clearly, as discerned from the reported R^2, our less preferred specification performs much worse in terms of explanatory power, and the statistical significance of estimated effects is lost in some cases. Nevertheless, these results appear to be generally reconcilable with estimates based on the logarithmic specifications.[49]

As mentioned in Section 5.1, several concerns about confounders in the analysis may be addressed by collapsing the 25 counties into larger geographical units. The estimates from these 'super-regions', also presented in Table C-1 in Appendix C, show that baseline results are robust to this alternative regional division.

In order to further check the robustness of our findings, we perform additional regressions not included in this version. First, to handle potential spatial heterogeneity of regions, we run regressions including the spatial lag of the treatment variable – i.e. the flu mortality in neighbor regions weighted by distance. Our point estimates of the treatment effect on poverty is hardly affected and also the non-finding for earnings persists. However, for capital income the standard error increases, reducing statistical significance. Second, since we find that earnings are unaffected by the flu we included earnings as a further control variable in the poverty and capital income regressions. Baseline results are completely unaffected by this modification. A third concern is whether the pandemic actually came as a surprise (especially the later waves). In order to address potential anticipation effects we estimate the effect including only the treatment of 1918. Baseline findings are not affected by these changes. Finally, to avoid the potential bias following from deaths caused by the flu being recorded as pneumonia cases – which, according to Figure 4.2, should not be a big problem – we also use information on influenza and

[49]Since all our dependent variables are weighted by total population and the number of inhabitants is directly affected by the flu, we also re-estimate our regressions dividing them by the population of 1917. This does not affect our results.

pneumonia from Statistics Sweden to derive a second version of our treatment variable. Results are not affected.

In conclusion, we find strong evidence for the pandemic having a positive impact on poverty in the medium term, and relatively strong evidence that capital incomes were negatively affected by the Spanish Flu. However, there is no evidence at all that earnings or labor productivity were affected by the pandemic. Placebo estimates are insignificant and close to zero suggesting that the common time trend assumption can be retained in all cases.

4.6.3 Reconciliation with the theoretical literature

Most of the predictions delivered by the theoretical model in Section 4.2 failed to be confirmed by our empirical analysis. We do not observe the expected immediate increase in wages – instead, the immediate impact on earnings is negative throughout, but the point estimate is small and nowhere near statistical significance at conventional levels. Besides, we observe a rapid decline in capital returns, even though these are predicted to increase by the same proportion as wages. Moreover, we observe an effect of the pandemic on poverty which goes far beyond the direct effect coming from dependents losing their breadwinners: on average, each influenza death resulted in four individuals moving into poorhouses. This finding suggests that poverty rates would have increased even if these dependents could be disregarded.

On the other hand, our results clearly suggest that more heavily affected counties experienced slower growth than the less affected ones in the aftermath of the pandemic. This appears to be the only prediction of the theoretical model which is not rejected by our empirical analysis. Since regional GDP is made up of returns to labor and capital, it is quite clear that the regional economies suffered a setback in economic activity during the pandemic (capital returns dropped and wages remained constant) which was reinforced in the years following the pandemic (capital returns dipped further whereas no further change in wages was observed).

In summary, our empirical evidence seems to support the notion of imbalance effects giving rise to slower growth in the post-influenza period, but it contrasts sharply with predictions of the model when it comes to the *immediate impact on GDP* and concerning the *distribution* of this impact between capital and labor. Below we provide a simple sketch of how our empirical findings can be reconciled with growth theory. Since the noted changes in the post-epidemic period are expected, it seems reasonable to focus on the immediate impact of the pandemic.

As emphasized by Boucekkine et al. (2007), it is commonly believed that population growth and population density have played an important role in the

transition from stagnation to modern growth, and the Spanish Flu pandemic possibly entailed a partial reversal of this process. In our case, the results suggest that there is a Solow neutral agglomeration externality in production, i.e. an effect of the density of the local market which tends to increase the marginal productivity of capital. This possibility has been analyzed in some detail by Acemoglu (2003) in the context of a growth model in which technological change may increase the productivity of either labor or capital. For illustrative purposes, we now consider a strongly simplified version of the production function for final goods, where we disregard the variables representing education u and human capital h:

$$Y = A\left(aK^{\phi} + bN^{\phi}\right)^{1/\phi} \tag{4.30}$$

In this model, a Solow neutral agglomeration externality would be represented by the parameter a, which also determines the share of capital returns in GDP.[50] Taking the derivative of GDP per capita with respect to population size, we get

$$\frac{\mathrm{d}y}{\mathrm{d}N} = \frac{1}{\phi}\left[\frac{\mathrm{d}a}{\mathrm{d}N} - \frac{\phi a}{N}\right] Ak^{\phi}\left(ak^{\phi} + b\right)^{\frac{1-\phi}{\phi}} \tag{4.31}$$

A sufficient condition for getting the desired result is that the agglomeration externality is positive and that

$$\frac{\mathrm{d}a}{\mathrm{d}N}\frac{N}{a} > \phi \tag{4.32}$$

Using this property, it is straightforward to show that the immediate impact of a pandemic will be

1. A reduction in GDP per capita.

2. A disproportionate reduction in capital returns, and thus

3. A redistribution of factor returns from capital to labor.

The literature on agglomeration economies has typically focused on local spillovers of knowledge and ideas, and it is less usual to find an agglomeration effect related to the marginal productivity of capital. However, seminal contributions by Dixit (1973) and Helsley and Strange (1991) have emphasized the role of scale economies and capital markets in production. Helsley and Strange (1991) show that if capital assets are specialized and immobile, their value in production will depend crucially on the population density of the region. Cingano and Schivardi (2004) identify a robust impact of specialization

[50]Note that a positive externality of labor, represented by the parameter b, would lead to a disproportionate decrease in wages, in contrast to our results. Since we are looking for a theoretical model that can explain our findings, we disregard this possibility in what follows.

and city size on firm-level total factor productivity. In a more recent contribution, Durlauf et al. (2008) analyze the performance of different growth theories under the assumption of model uncertainty, and conclude that externalities in physical capital obtain a much higher explanatory power (and thus inclusion probability) than many variables associated with standard growth theories.

As alternative explanations for our somewhat counterintuitive results would be *selective mortality* and *scarring*. Selective mortality could give rise to our results if physically fit individuals had higher mortality than the rest. Such a selection would seem unlikely in general, but since the 1918 influenza was particularly prevalent among young and healthy adults, it is not completely implausible. However, even though this selective mortality were strong enough to lead to a reduction in GDP per capita (despite capital deepening) it cannot explain the change in the shares of labor and capital incomes in GDP. To the extent that scarring occurs, i.e. that influenza survivors experience a deterioration in their health, we would also observe a reduction in labor productivity which could offset and possibly exceed the effect of capital deepening. But again, this would not lead to capital returns being particularly hard hit, and besides, the fact that our results are robust to the inclusion of infection rates strongly suggests that scarring does not drive our results.

4.7 Discussion

It has been argued that regional differences in exposure to the 1918 influenza pandemic were largely random. If this holds to be true, these regional patterns in mortality rates can be exploited to estimate the effects of a substantial health shock to the economy. Such an exercise has the potential to shed light on at least three important issues. Firstly, providing an estimate of the actual economic consequences of the 1918 pandemic. Secondly, giving us an idea of the possible effects of current and future pandemics on the performance of the economy. Thirdly, we might be able to say something in general about the functioning of the economy, and how labor supply shocks are transmitted through the system.

We have shown that the Lucas (1988) model of endogenous growth delivers a set of clear predictions of how an economy can be expected to react to a epidemic of this kind. The immediate effect will be an increase in wages and capital returns, and a reduction in interest rates. This effect is a direct consequence of capital deepening and it will normally not be accommodated by workers moving from education into production. For the medium term, the model instead predicts negative imbalance effects on growth as long as the ratio of physical capital to human capital remains above the long-run equilibrium

value.

Our study finds no evidence against the assumption that the epidemic was a largely random shock to Swedish regions. The common time trend assumption appears to be satisfied for all variables, and we also fail to identify a socioeconomic gradient in the incidence of the epidemic. Besides, since influenza incidence and mortality tend to follow the same spatial patterns in general, it is reassuring that our main results are robust to the inclusion of variables capturing different aspects of influenza incidence. Thus, it is our tentative conclusion that differences in excess mortality rates across regions are largely exogenous.

Our main findings are generally very robust. For capital incomes, we find that the pandemic had a strong negative impact, and this impact appears to have been a combination of immediate and medium-term responses. According to our estimates, the highest quartile (with respect to influenza mortality) experienced a drop of 14 per cent during the pandemic and an additional 12 per cent afterwards. For earnings, on the other hand, we are unable to detect any effect either during or after the pandemic. For poverty, finally, we find a strong and positive effect, which seems to have appeared only once the epidemic had receded in 1920. For this variable, the top quartile suffered an increase in poverty by 23 per cent compared to the bottom quartile.

Strong as these results may seem, they do not fit very well with the most popular macroeconomic models. On the one hand, we do get the result that heavily affected counties had lower growth rates after the epidemic. On the other hand, our findings that earnings were unaffected and capital incomes dropped is much more difficult to explain. Likewise, our finding that poverty rates increased is also difficult to reconcile with the increased scarcity of labor.

If our empirical results reflect the true causal effect of the pandemic on the economy, they give some useful insights into how the regional economy should be modeled. Clearly, there is a redistribution between capital and labor taking place, and this redistribution in turn suggests that the marginal product of capital is subject to an agglomeration externality. There is a rich literature in regional science on how such externalities might arise, and some previous empirical research supports the notion of an externality related to physical capital. This clearly seems to be an important issue for further research to understand the economic consequences following major health- and labor supply shocks.

Before reverse the causal chain in the next Chapter we briefly summarize our findings. The analysis of Sweden showed that there are large effects of the Spanish Flu on capital income and poverty, while wages were not affected. Overall this would suggest a decrease in GDP per capita, as shown in our

theoretical model in this Chapter. However, in the Chapter 2 we do not find these effects in all countries. There could be two factors explaining this: the nature of the disease and the analyzed countries. Compared to the Spanish Flu, HIV is a much more slow-moving and persistent disease. This gives more time to react and stabilize growth. In terms of analyzed countries it might well be that foreign aid funds helped African countries to deal with HIV, while this was not possible for the Spanish Flu because it happened much faster and affected the whole world. Moreover, Bloom and Mahal (1997b) argue that African countries have many "unused resources" in terms of labor force, which might explain the non-finding in some African countries. Finally, one of the main problems in Africa remains a high birth rate and our results in Chapter 3 show that that HIV did surely not decrease fertility.

CHAPTER

5

Sickness Absence and the Business Cycle: Moral Hazard, Labor Force Composition or Neither?

5.1 Motivation

After we analyzed how the economy is affected by infectious diseases, we reverse the causal channel in this Chapter and investigate how the business cycle affects the spread of infectious diseases. It is a stylized fact that sickness absence is procyclical. In the literature this has mainly been documented by showing a negative correlation between sickness absence and the unemployment rate: During periods of low unemployment sickness absence is high and vice-versa.

One frequently cited explanation for the phenomena of procyclical absence rates are *labor force composition* effects (see for instance Leigh 1985). The explanation is based on individuals who exhibit a higher tendency to be sick. Due to this health condition, they will be unemployed more often, especially during recessions. During booms, however, they will also be able to find jobs, but will cause higher sickness absence and thus procyclical absence rates, due to their worse health.

While health surely plays an important role for reporting sick leave, there

might be also other relevant factors. Many scholars report that incentives matter for sickness absence as well. Too high sick pay might lead to moral hazard and overreporting of sickness (see for instance Johansson and Palme 2005, Puhani and Sonderhof 2010, Ziebarth and Karlsson 2010).

These opportunity costs of sickness absence might change over the business cycle leading to procyclical absence rates. It could be argued that individuals have lower costs of absence when unemployment is low, since they are more difficult to replace. On the other hand, their fear of job loss might be higher during a recession inducing them to stay at home only when it is absolutely necessary. A similar argument was put forward by Shapiro and Stiglitz (1984), who suggest a model where unemployment works as a disciplining device, reducing workers' *moral hazard* and leading to increased effort during an economic recession. This higher effort could lead to lower sickness absence and could explain procyclical sickness absence.

Several empirical findings support this view. For instance, Ichino and Riphahn (2005) look at a large Italian bank and find that once the 12-week probation period of newly employed individuals has elapsed, the absence rate increases sharply due to greater job protection and a lower firing risk after the probation period. Lindbeck et al. (2006), using Swedish data, also look at job security and find that more job security leads to higher absence. Another study by Bratberg and Monstad (2012), where the authors employ a negative financial shock that hit specific employers as an instrument, can also confirm this finding.

While job security might have an impact on shirking and thus sickness absence, the (local) unemployment rate might not necessarily be a good proxy for job security. In particular, workers might be mobile and/or might not observe the (local) unemployment rate. Therefore, we suggest a new explanation driven by *presenteeism* (working while sick) *and infections*. During expansive periods the workload of employees is higher than usual. For instance Millard et al. (1997) report higher average hours during periods of expansion. Some employees who get sick during this period might still decide to come to work (possibly with the help of some medicine) because their tasks cannot be taken over by somebody else due to increased work specialization. If the physical distress is caused by a contagious disease, appearance at work might lead to infection of their co-workers and cause procyclical sickness absence.

Presenteeism has been subject to much medical research in the last decade. Aronsson et al. (2000) conducted a survey and found that around 50% of the employees went to work at least once within the last year despite feeling sick. Another study by Goetzel et al. (2004) analyzed the costs of the ten most costly diseases for companies. Using data from the US, the authors find that

the costs for presenteeism exceed the costs for absenteeism by far.

Several studies analyze what causes presenteeism. Demerouti et al. (2009) for instance report that higher job demand – like more projects and more time pressure – leads to more presenteeism. Moreover, Aronsson et al. (2000) and Aronsson and Gustafsson (2005) find that presenteeism is higher if tasks are more specialized and cannot be taken over by co-workers and for individuals with more time pressure. Hansen and Andersen (2008) also provide evidence that time pressure leads to more presenteeism.

A recent paper by Barmby and Larguem (2009) documents infections at the workplace and provides further support for the suggested mechanism. They are the first to combine absence literature with the epidemiology of infectious diseases. Using daily absence data of a factory in the UK, the authors find significant effects of workers' absence on the absence of their peers. Even though they have no data on whether the absence was caused by infectious diseases, it is suggested that infectious diseases are the main force behind their findings.

The empirical literature aimed to test the two explanations based on composition effects and moral hazard by analyzing how the unemployment rate is related to sickness absence.[51] Leigh (1985) is the first to study how unemployment affects sickness absence. Their findings, based on individual level data from the US, suggest that both composition effects and moral hazard play an important role. Since the effect for the whole labor force (employed and unemployed) is generally larger (in absolute values) than for employed individuals only, they conclude that composition effects are an issue. Moreover, they interpret the negative correlation among employed individuals as an indication for the moral hazard hypothesis in the spirit of Shapiro and Stiglitz (1984). However, it is not clear whether the fear of job loss drives the finding of Leigh (1985). As we will show below, our explanation based on presenteeism and infectious creates similar predictions about sickness absence and the business cycle.

The fact that sickness absence is procyclical is also confirmed by Johansson and Palme (1996). Using Swedish data, they also find that the unemployment rate has a significant negative influence on sickness absence. However, they do not investigate this finding further and don't discuss possible explanations.

A more recent paper by Scoppa and Vuri (2012) also investigates whether moral hazard is the driving force behind procyclical sickness absence. Using

[51] There is a related literature on procyclical mortality started by Ruhm (2000). However, in a recent paper Miller et al. (2009) suggest that the mechanism involves more road accidents due to more traffic during periods of economic growth. Relating this to the workplace, Davies et al. (2009) looks at occupational injuries and also finds that minor injuries are procyclical, while they find no effects for major injuries. However, workplace accidents are not procyclical in our data and thus cannot be the driving mechanism. We therefore disregard this channel.

data from Italy, they find that workers in small firms react more to local unemployment than workers in larger firms, while employees in the public sector show no reaction at all. Again they argue that this is because workers in small firms have a greater fear of job loss, since they are less protected in terms of employment protection legislation. However, this result could also be driven by the fact that smaller firms have greater difficulties reacting to the business cycle and hiring new workers and thus the exerted pressure on existing workers is even larger during economic booms.

A further investigation of the two explanations for procyclical sickness absence is presented by Arai and Thoursie (2005). They suggest that the two hypotheses generate opposite predictions on the correlation between sick rates and the share of temporary contracts. With many temporary contracts job insecurity is high and thus absence should be low and therefore there should be a negative relationship. On the other hand temporary contracts are for marginal workers, which suggests a positive relationship between temporary contracts and absence. The authors rely on aggregated industry region data from labor contract data and absence rates in Sweden and find a negative correlation between the two measures. Thus they conclude that the moral hazard effect dominates.

Audas and Goddard (2001) use US data and – relying on time-series econometrics – estimate a cointegration model looking both at the labor supply and demand side. On the labor supply side they find a negative relationship between sickness absence and unemployment and are thus able to confirm previous findings. On the demand side, theory would suggest that during times of high product demand sickness absence is more costly. The authors confirm the theoretical predictions since they find that monthly industry production is negatively related to sickness absence.

Using data from Norway, Askildsen et al. (2005) also study how the unemployment rate affects sickness absence. In Norway a doctor's certificate is not needed for the first 14 days of absence. Therefore, the authors are only able to analyze absence spells longer than 14 days. They use individual-level panel data and find that the negative relationship between absence and unemployment is more pronounced for the stable labor force workers ("insiders"), while the effect is weaker over the whole labor force. Therefore they can rule out a pure composition effect and suggest that the moral hazard hypothesis drives their results.

A similar result is found in Fahr and Frick (2007) using data from Germany. The authors use state level data from 1991-2004 and find that older workers react most to the unemployment rate. They suggest that this is caused by the higher opportunity costs of losing their jobs. Thus they argue that the results

are driven by moral hazard effects.

Dyrstad and Ose (2002) using data from Norway can also confirm this result. They distinguish between composition and moral hazard effects using the level of unemployment. The authors suggest that moral hazard effects dominate if the level unemployment is high, since relatively few individuals with bad health are in the labor market. On the other hand, if unemployment is low, composition effects dominate, while the fear of job loss is negligible. The authors a use smooth transition regressions such that high and low levels of unemployment arise endogenously and find that for long-term sickness discipline effects matter, while for short term sickness both effects are present for men, while neither is present for women.

Nordberg and Røed (2009) also look at the relationship between unemployment and sickness absence using individual level data from Norway for long-term sickness (15 days or more). They also find that absence is procyclical and suggest that the finding is due to the moral hazard of employees. Their empirical model ensures that the results are not driven by the composition of the labor force, although in a second step they show that sorting also plays an important role. Thus they find empirical evidence for both explanations.

Summing up, many scholars find that absence rates are procyclical. Moreover, most findings seem to suggest that the fear of job loss causes this finding. However, the fact that most of the analyzed countries provide quite generous unemployment benefits raises doubts whether this mechanism is an adequate explanation. Furthermore, for long-term sickness absence (longer than two weeks) it seems difficult to argue that moral hazard is the main driving force (as suggested in Askildsen et al. 2005). Moreover, as seen in Audas and Goddard (2001) labor demand is countercyclical. In particular, sickness absence is more costly for firms during peak times. Therefore it is surprising that firms are not able to reduce sickness absence by exerting more pressure and higher monitoring during economic booms in order to reduce absenteeism. Hence, we suggest an alternative channel, where the additional sickness absence during economic booms is caused by infections. We argue that most of the empirical papers above do not directly test the model suggested by Shapiro and Stiglitz (1984), since it is not clear whether the fear of job loss provokes lower sickness absence during times of high unemployment. As we will show in our theoretical model below, our explanation is based on the fact that sickness absence is initially *lower* during periods of economic booms, due to higher monitoring and more pressure. In fact the additional pressure will incentivize sick individuals to go to work even though they are sick. For infectious diseases this will lead to infections of coworkers and thus cause procyclical sickness absence.

In a second step, we use data from 112 (out of 145) public health insurance

funds in Germany to test this model. The data is aggregated at state level and covers the years 1999 to 2010. Due to the panel data structure, we are able to control for unobserved heterogeneity between German counties. In terms of data our contribution is that we can analyze the cause of sickness, which is provided in different categories. Therefore we are able to disentangle the correlation between unemployment and overall sickness behavior from infectious diseases.

The results show that sickness absence for infectious diseases shows a more pronounced relationship with the business cycle than sickness absence for non-infectious diseases. This provides support for the mechanism suggested above. However, for diseases of the respiratory system (such as the flu) we find no relation to the business cycle at all. We suggest that this might be related to the fact that for these diseases too many infections do not happen at the workplace.

Our analysis extends the previous literature in several ways. Firstly, we provide a new explanation for the negative correlation between sick leave and the unemployment rate. Secondly, we are able to test different theories on this negative correlation due to the cause of sickness absence. German sickness funds are obliged by law to collect data on their insured individuals. Therefore, our data is more reliable than results based on survey data and the measurement error in this data is lower than in other studies.

This Chapter is structured as follows. In Section 5.2, we provide a short theoretical model. In Section 5.3, we present an introduction to the health care system and sickness payments in Germany. Section 5.4 provides the data and empirical specification. Finally, we present the results in Section 5.5 and discuss them in Section 5.6.

5.2 A model of sickness absence with infections

We assume an economy with two states ξ_t, booming ($\xi_t = h$) and in recession ($\xi_t = l$). Time is discrete in our model and we assume that one time interval is equal to one week. As we will discuss in more detail below, this assumption is handy for dealing with sickness and recovery. The states of the economy will vary in cycles to resemble the business cycle, meaning that we will have many weeks forming a long time interval of a booming economy followed by a recession. Whether the economy is booming or in recession affects the income of working individuals $y(\xi_t)$, with $2 > y(h) > 1 > y(l) > 0$. Income for non-working individuals, i.e. individuals on sick leave, is normalized to zero. The reason for income when working being higher than zero is to ensure that

individuals go to work when they are healthy.[52] The higher income in the good state of the economy can be justified in various ways. One could think of an incentive contract where the worker gets a higher income when the sales of the company increase. Non-monetary costs and benefits could also play a role. There are probably more projects in a company during a favorable economic period, which might lead to private benefits for the worker through feeling needed in the company. Moreover, since income on sick leave is normalized to zero, it is, ceteris paribus, more costly to go on sick leave when the economy is in a good state. More projects might lead to more upcoming deadlines and usually the time shortly before deadlines is characterized by high pressure and therefore sick leave might be an option only for very serious sicknesses. Another related explanation is that tasks are usually quite specialized and therefore the workload will not decrease when calling in sick, making sick leave more costly during periods of economic growth. Finally, it seems likely that labor demand is higher in good states and that the supervisor exerts some pressure on the employee making sick leave more costly.

We denote sickness with the variable θ_t, with $\theta_t = 1$ in case of sickness and $\theta_t = 0$ otherwise. Working individuals face some costs of working while sick equal to $c_i\theta_t$. These costs differ between individuals. Costs will be low ($c_i = 1$) for a proportion q of the population and high ($c_i = 2$) for the remaining fraction $1 - q$.[53] We need both groups for our model to ensure that some individuals come to work and infect others, while other individuals stay at home and cause an overall higher sickness absence. Finally, we assume the following utility function:

$$U_{it} = y(\xi_t) - c_i\theta_t. \tag{5.1}$$

All individuals have information on their current sickness status θ_t and compare costs and benefits of working as opposed to sick leave. Each week the workers decide on whether to go to work or stay on sick leave, given their income from working $y(\xi_t)$ and sickness costs $c_i\theta_t$.

Given some time-varying sickness probability p_t, which will be discussed more in detail below, we assume a sufficiently large population (normalized to one) so that p_t also represents the fraction of the population falling sick. Table 5.1 below displays the proportions of individuals who will be working and on sick leave in the various states of the economy.

[52] In countries where sick pay is equal to the salary (as it is the case also in Germany) this can be justified by the fact that calling in sick when healthy might lead to discovery and job loss, especially if this occurs frequently. Moreover, individuals who claim that they are sick when they are not might feel guilty about this behavior which might reduce their non-monetary income.

[53] Continuous sickness costs with a threshold above which sickness costs are too high to go to work in the good state would produce similar results.

No matter what the current state of the economy, healthy individuals $(1 - p_t)$ will go to work since they earn a positive income, while they have to bear no costs for sickness. For non-healthy individuals (p_t) work participation depends on the current state of the economy. If the economy is in a low state they only earn $y(l) < 1$ and face sickness costs of one or two respectively. In any case these costs exceed income and therefore both groups decide to go on sick leave. In the good state of the economy opportunity costs change since working individuals now earn a higher income $y(h) > 1$. For individuals with low sickness costs $(p_t q)$ this will suffice to prefer work to sick leave, while individuals with high sickness costs $(p_t(1 - q))$ still prefer sick leave.

Table 5.1: Proportion of Population Working and on Sick Leave

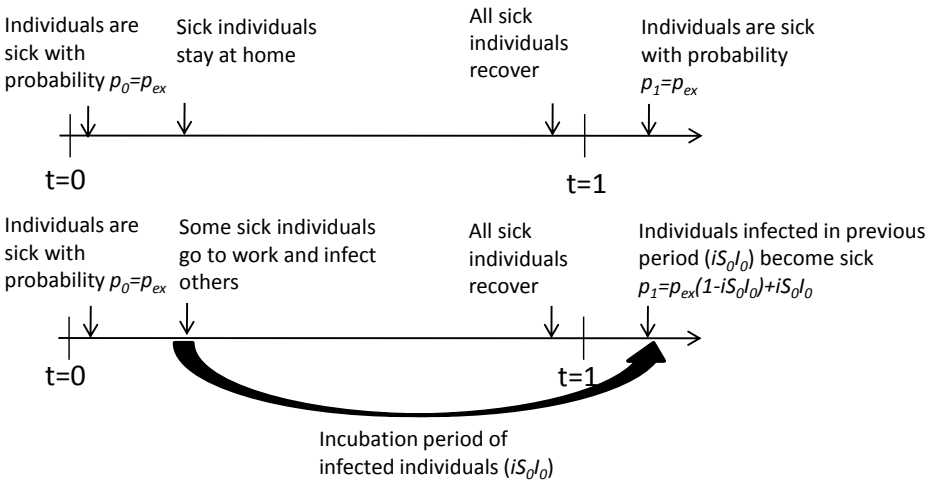
	working, $\theta_t = 0$	working, $\theta_t = 1$	sick leave
$\xi_t = l$	$1 - p_t$	0	p_t
$\xi_t = h$	$1 - p_t$	$p_t q$	$p_t(1 - q)$

The Table displays the fractions which are working, sick $(\theta_t = 1)$ and on sick leave in the two states of the economy, namely the booming $(\xi_t = h)$ and the recession state $(\xi_t = l)$.

The final ingredient for our model are infections and recovery. The timing of infections and recovery is as follows: At the beginning of every week t there is some probability p_t for individuals to fall sick. Depending on their sickness status θ_t and the current state of the economy ξ_t, sick individuals choose whether to work or not. Finally, at the end of the week all previously sick individuals recover, thus one week t resembles the time needed for recovery. We assume that infections occur at the workplace, where healthy individuals (susceptible) will be infected by their sick coworkers. The infected co-workers will not fall sick immediately, but will fall sick only after an incubation time equal to one week as well. For this reason the infections will affect the sickness probability in the next period p_{t+1}. Figure 5.1 below summarizes these dynamics in the two states of the economy.

During a period of recession $(\xi_t = l)$ there will be no sick individuals at work and therefore no infections occur and the sickness rate $(p_t = p_{ex})$ does not change from one period to the next. All individuals recover at the end of the week. Subsequently, a new week starts where the sickness probability is again equal to an exogenously fixed parameter. If the economy is in a good state however, some fraction of sick individuals $p_t q$ will come to work and infect their peers at the workplace. Again, by the end of the week all previously sick individuals recover. However, the individuals who have just been infected do not recover and fall sick in the following week and increase the fraction of sick individuals. Finally, also in this state we have an exogenous sickness

Figure 5.1: Timeline



probability p_{ex} – which is restricted to non-infected individuals.

Summing up, we will have the following dynamics for the probability of falling sick:

$$p_{t+1} = f(p_t) = p_{ex}(1 - iS_tI_t) + iS_tI_t \tag{5.2}$$
$$= \begin{cases} p_{ex} & \text{if } \xi_t = l \\ p_{ex}(1 - ip_tq(1 - p_t)) + ip_tq(1 - p_t) & \text{if } \xi_t = h \end{cases}$$

The dynamics above follow from adapting the SIS (susceptible-infected-susceptible) endemic model to our setup. Individuals in the model are either infected (I) or susceptible (S). The infection rate at any point in time reads as: iS_tI_t, where i is a parameter measuring how easily the disease is transmitted between individuals and S_t and I_t are the fraction of susceptible and infected individuals respectively. The model assumes that all persons in the population meet once and infections occur at contact. The transmission parameter i can be interpreted as the percentage of meetings resulting in an infection. Thus if $i = 5$, then 5% of all meetings result in infections. Replacing S_t and I_t with the respective fractions from columns one and two in Table 5.1 above and adding some exogenous sickness rate p_{ex} we obtain the dynamics above (5.2). In the booming state, the infection rate might change at each point in time, with the resulting change in dynamics. However, depending on the parameter choice an equilibrium will be reached where in each period the number of infections equals the number of recoveries.[54]

[54]In this kind of model a recovery rate also usually appears. However, for simplicity we assumed that all individuals recover at the end of one period and thus we normalized the recovery rate to one.

The SIS model is the classic framework of mathematically analyzing infectious diseases and was first discussed in the medical literature by Ross (1916) and Kermack and McKendrick (1927). In the basic models individuals are initially susceptible and become infected afterwards. After a certain time they recover and become susceptible again. These models were used to study epidemics in order to identify different dynamics and find conditions for outbreaks of epidemics and how epidemics can be eradicated through vaccinations. The models were developed further by many scholars, who introduced time varying transmission parameters, stochastic infections, nonlinear infection probabilities and added more population groups, such as recovered individuals who develop immunity (see for instance Philipson 2000, Capasso 2008, for recent surveys). In the economic literature one can also find applications of the SIS model. For instance, Gersovitz and Hammer (2004) look at the general problem of the social planner and how to prevent the spread of diseases. Francis (2004), on the other hand, studies the special case of optimal subsidies for flu vaccination. Finally, Laxminarayan and Brown (2001) investigate the optimal use of antibiotic resistance.[55]

We proceed by calculating the equilibria of the sequence for the booming economy. A fixpoint will be reached if $p_{t+1} = p_t$. Given the structure of the sequence above the fixpoints in this case will equal:

$$p^*_{\pm} = \frac{1}{2iq}\left(-1 + iq \pm \sqrt{1 - 2iq + 4iqp_{ex} + i^2q^2}\right). \tag{5.3}$$

Figure 5.2 below draws graphs for three exemplary combinations of $\{i, q, p_{ex}\}$. Fixpoints are obtained when the 45 degree line is crossed.

Finally, we obtain the following proposition stating the conditions for stable equilibria.

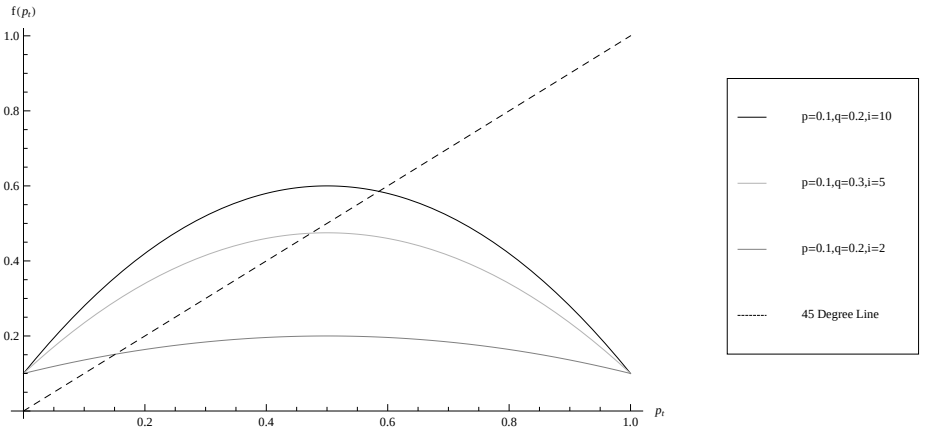
Proposition 1 *The infection rate will reach an equilibrium with higher sickness absence during periods of economic booms and lower sickness absence in recessions*

$$\begin{aligned} &\textit{if}\left(p_{ex} = 0 \ \textit{and}\ \frac{1}{q} < i < \frac{3}{q}\right) \ \textit{or} \\ &\left(0 < p_{ex} < \frac{1-q^2}{1+2q} \ \textit{and}\ \frac{1-q}{(1-p_{ex})^2 - q(1-p_{ex})} < i < \frac{2\sqrt{1 - p_{ex} + p_{ex}^2} + 1 - 2p_{ex}}{q(1-p_{ex})}\right) \end{aligned} \tag{5.4}$$

Proof. See Appendix D.

[55] Similar models are also used in the economics of innovation literature in order to model new innovations and their diffusion (see Dosi and Nelson 2010, for a recent survey).

Figure 5.2: Cobweb Diagram



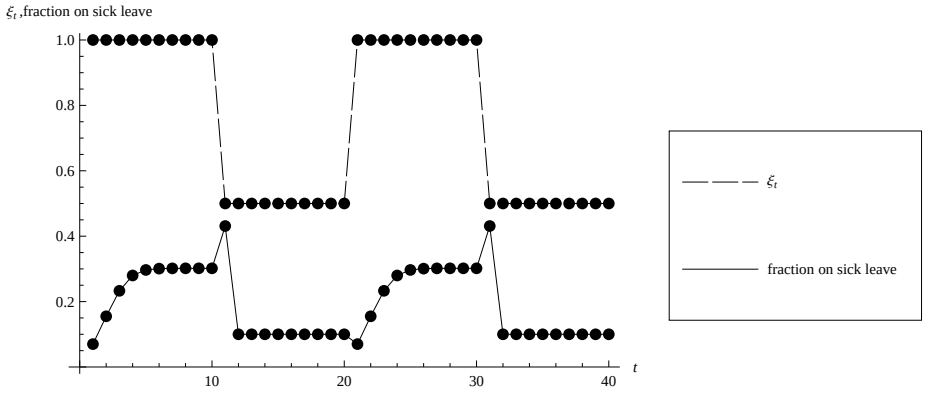
Intuitively, if the transmission parameter i or and/or the fraction of sick people going to work q are too large the fixpoints are not stable. Moreover, an excessively high exogenous sickness rate p_{ex} will increase sickness absence during recessions and therefore decrease the procyclical association of sickness absence. Furthermore, if the transmission parameter i is to small the disease will die out (if $p_{ex} = 0$) or the sickness absence will be smaller during economic booms (if $p_{ex} > 0$). With respect to the fraction of sick people going to work q the analysis is more complicated. On the one hand, if many sick individuals go to work there will be no infections. On the other hand, many individuals going to work means nobody stays at home and therefore sickness absence cannot increase.

Summing up, our model shows that infections at the workplace can lead to higher sick leave during periods of economic booms. Figure 5.3 below graphs this development. The dashed line represents the booms and recessions in the economy (ξ_t) while the solid line represents the fraction of individuals on sick leave. Initially, the good state of the economy persists, leading to infections and an increase in sickness absence. As the economy turns into a recession, sickness increases initially since now all sick individuals are on sick leave while before only individuals with high sickness costs were on sick leave. Afterwards no infections occur and sick leave decreases rapidly. At the end of the recession sick leave shortly decreases again, since during the boom only individuals with high sickness costs go on sick leave.

We will test the model by looking at the relationship between aggregate sick leave data and the business cycle in Germany. Since the mechanism described above will only appear for infectious diseases, we expect larger point estimates for infectious diseases as compared to non-infectious diseases. In fact, our model predicts a counter-cyclical relationship for non-infectious diseases. Since

Figure 5.3: Development over the Business Cycle

for $p_{ex} = 0.1$ and $q = 0.3$, $i = 5$; $\xi_l = 0.5$, $\xi_h = 1$



more people come to work even though they are sick during booms, and no mechanism of infections exists, we should find lower sickness absence during booms for non-infectious diseases. We proceed by describing the institutional setting in Germany, before we present the data and the results.

5.3 Institutional Background

5.3.1 Health Insurance in Germany

In Germany, both private and public health-care providers coexist and nearly all inhabitants are covered by health insurance. Around 90% of individuals are with public health care providers (German Federal Ministry of Health 2012). Individuals with an annual salary above 45,900 € in 2003 and self-employed can choose between public and private health insurance, while below this threshold public health insurance is mandatory for employees.[56] Students and unemployed form special groups. However, they are also mostly insured with public providers. Health insurance is financed by a premium that is paid as a share of salary, and non-working spouses and children are covered by the family health care insurance of the bread-winner.

In the public sector there were a total of 145 health insurance funds in September 2012 (German Federal Ministry of Health 2012). Since 1996, individuals in Germany are free in the selection of their statutory health provider, while before they were allocated to the different funds, based on their occupation or industry. The packages offered by the different health insurance funds are very similar, because basic health provision is the same at federal level and ensures that most health expenses are covered. Moreover, there are additional

[56]This threshold changes every year, depending on the average salary growth (current value in 2013 50,850 €). For more information see §6 Sozialgesetzbuch.

packages that might be chosen and paid separately, which give special coverage for eyeglasses, dental services, etc. The health insurance premium does not depend on the individual health status and future health risks. Therefore, the basic product offered is quite similar and competition happens mainly at a price level.

Public health insurance is financed by mandatory payroll deductions of 15.5% from the individual payroll. Payments are split between employer and employees with a share of 7.3% and 8.2% of gross wage respectively. These funds are collected and used to finance the public insurance providers. If the provider has a negative balance or a surplus, an additional deduction or premium may result. Health insurance providers are competing for the insured with the lowest risk of bad health and high payments, while at the same time they are not allowed to reject individuals because of their bad health. However, a risk equalization scheme is in place, where funds with a good risk insurance pool have to contribute, while the fund compensates providers with more bad risk pools.

5.3.2 Sick pay and monitoring

Compared to other countries, the sick pay system in Germany is quite generous. In terms of rights, employees get fully reimbursed for their foregone wage during the first six weeks of sickness.[57] After this initial period reimbursement goes down to 80% and the health insurance fund has to bear the costs, while the first six weeks are paid by the employer.

On the other hand, employees have the duty to provide a doctor's certificate as of their fourth day of sickness or earlier if asked so by the employer. The doctor's certificate provides the primary monitoring. Moreover, there is an independent medical service, which performs additional monitoring. It is usually consulted by the health insurance funds or the employer, when doubts about work absence exist. Doubts may arise due to frequent sickness of individuals or certain doctors issuing an unusually high number of certificates. The medical service employs independent doctors who may examine the documents of the patient or even the patients themselves. In fact in 2011 about 1.5 million cases were examined by the medical service. (Medizinischer Dienst der Krankenversicherung (MDK) 2012)

In Germany sick pay is available both for unemployed and for employed individuals. Unemployed individuals are supposed to search actively for a new job. Since sickness provides an additional hurdle to job search, unemployed individuals with sickness absence are entitled to sick pay. The first six weeks

[57] Reimbursement was dropped to 80% for these six weeks from 1996 to 1998. See Puhani and Sonderhof (2010) for further details.

of sickness absence are paid by the unemployment agency. One might argue that the unemployed have no incentive to call in sick, since the sick pay equals their unemployment benefits. However, during the first six to twelve months of unemployment – depending on the age and the previous employment of the individual – the benefits obtained from the unemployment agency depend on the previous salary and – depending on whether the recipients have children – amount to around 60 or 67% of the previous salary. After this period they are reduced to basic social security, which is independent from previous salary. Therefore, a period of sickness (notified to the authorities) results in a longer time interval of higher unemployment benefits.

5.4 Data and Empirical Specification

5.4.1 Data

German insurance providers are grouped as follows. Out of the 145 health insurance funds 11 funds form the AOK (Allgemeine Ortskrankenkassen), with a total of 11.3 million members. The category with the largest number of members is the VdEk (Verein der Ersatzkassen) with 13.5 million members and 6 health insurance funds – noteworthy health insurance funds with many members in this category include Barmer and the TK (Techniker Krankenkasse). Then come the BKK (Betriebskrankenkassen), with a total of 112 health insurance funds and 6.2 million members. Other insurance funds are the IKK (Innungskrankenkassen) with 5.4 million members, the Knappschaft with 1.8 million members and the LKK (Landwirtschaftliche Krankenkasse) with 0.8 million (German Federal Ministry of Health 2012). The sickness funds are competing with each other for customers, even though they are linked by these broader categories. However, among the AOK health insurance funds competition is somehow limited due to their regional specialization.

Our data consists of BKK members.[58] Our primary outcome variable is sickness absence aggregated at state level. The data was collected from different yearly issues of the health report (Gesundheitsreport), a yearly newsletter published by BKK. Since there are 112 BKK funds offering different packages and different prices there is no obvious selection into a specific fund. The result is a twelve year panel (ranging from 1999 to 2011, with missing data for 2006), which has data on sickness absence aggregated at the state level (NUTS 1) of individuals who are paying-members of the BKK. What should be noted is that only reported cases are included. For sicknesses that lasts three days or less it is not mandatory to provide a doctor's certificate and thus there will be

[58]Data from other insurance providers did not distinguish the cause of sickness absence at state level.

an underreporting of these cases.

The paying members consist of non-self-employed individuals from the work force with an income below the threshold described at the beginning of Section 5.3.1. Moreover, short-term unemployed individuals are also included in the sample.[59] Unfortunately, we cannot distinguish between sickness absence of employed and (short-term) unemployed individuals. Therefore, we are not able to analyze to which degree sickness absence is driven by labor force composition effects.

We add additional data from various sources. Additional data on demographics of the members was collected by contacting representatives from BKK. Moreover, we obtained additional demographic data from the German Employment Agency (Statistik der Bundesagentur für Arbeit 2013). This data was matched with economic data and health at state level, consisting of the unemployment and death rate. The data for these two variables was obtained from the German Federal Statistical Office (Statistisches Bundesamt 2013a,b).

Sickness absence is categorized into different categories based on the International Classification of Diseases (ICD). From these categories we form 3 groups in order to test the hypothesis from our theoretical model, i.e., whether the relationship between sickness absence and the business cycle differs between infectious and non-infectious diseases. More specifically, we expect a higher coefficient (in absolute terms) for infectious diseases. The three groups thus consist of non-infectious diseases, infectious diseases and infectious diseases of the respiratory system. We separate these last two groups, since the model parameters for infection and transmission might differ for different kinds of infectious diseases. Respiratory diseases are more infectious, which might affect the transmission parameter i. Moreover, the proportion of individuals going to work sick q and the exogenous sickness probability p_{ex} might also differ. Finally, diseases of the respiratory system include both infectious and non-infectious diseases. Even though most cases of sickness absence in this category are due to infectious diseases (as shown below) it might be preferable to have a separate group for this mixed category. The details on how these three groups are formed from the ICD categories are available in Table D-1 in Appendix D.[60]

[59]The main distinction between paying members and insured individuals is that the members are paying regular fees as a percentage of their salary and in this way they obtain insurance for themselves and their family members, such as children and non-working spouses. Since our focus is on sickness absence of employed individuals we are mainly interested in the outcomes for paying members.

[60]Data after 2000 uses the categories displayed in the Table above, based on the 10th revision of the International Classification of Disease (ICD-10). Earlier data (1999-2000) relies on the older ICD-9 classification. However, the year dummies will capture this recategorization since they affect all counties equally. Moreover, following the fact sheet of the American Medical Association (2012) the main difference between the 9th and 10th revision

Before we come to the empirical specification and the results we will make the categorization more intuitive and easier to follow by providing some examples for the various categories. Most days at work are missed due to back pain. For this distress in 2010 around 450,000 absence spells were started, which resulted in a total absence of around 7 Million days. Back pain is categorized with M54 and forms about half of the diseases of the musculoskeletal system (category M) as defined in Table D-1. Other common conditions from the same category are herniated disc (M51), knee (M23) and shoulder (M75) problems. The category of diseases of the musculoskeletal system is the main entry in the group of the non-infectious diseases by far. The second largest category is composed of injuries and poisoning (S-T), with about 520,000 cases in 2010, with only one third resulting from workplace injuries. Category R stands for other symptoms such as malaise and fatigue (R53), abdominal and pelvic pain (R10), dizziness and giddiness (R42) and headache (R51). This category is also included in the group of non-infectious diseases. The only categories not included in this group are certain infectious and parasitic diseases (categories A to B) and diseases of the respiratory system (category J). In terms of infectious diseases (categories A to B) the most common one is an "infectious gastroenteritis" with close to 300,000 cases and 1,3 Million days and various viral infections (B34) with 200,000 infections and 1 Million days. All the entries from categories A to B form our second group. Respiratory diseases (category J) form the last group. The most common sickness is an "acute upper respiratory infection" (J06). However, this category also includes chronic diseases of the respiratory system (J40-J47) making up 15% of sickness absence cases due to diseases of the respiratory system.

Summing up, we will form three groups, where the first group includes non-infectious diseases. The second group consists of various infectious diseases (Categories A to B), while the third category is made up of diseases of the respiratory system (J).

Table 5.2 displays the summary statistics. In total we have 192 observations formed by 16 German states observed in 12 years (1999-2011 with 2006 missing). The first three rows show the summary statistics for the different disease groups. Non-infectious diseases show the highest sickness absence with sickness absence 70.4 spells per 100 members per year. The second group is

is a more detailed categorization. Due to the fact that our data is in broad categories in any case, there is little difference between the two categories for our analysis. The only main difference is that in earlier data categories G and H were previously aggregated to one category. We also performed additional regressions with ICD-10 data only (and thus using only data after 2000), however this did not affect the results. Another minor issue is that the classifications are updated every year and therefore there exist slight differences. However, again at our broad level this should not be a problem, and since the updates happen simultaneously in the whole country our time dummies will most likely capture arising differences.

formed by infectious diseases, with 8.9 sickness spells started. Finally, for respiratory diseases 32 sickness spells were started each year. Moreover, there appear no large differences in terms of gender. Only for respiratory diseases we find that women are sick more often, while for the other two groups no statistically significant differences appear. The number of members measures 330,000 on average in every state in Germany. Moreover, about half of the members are female. As is standard in this literature, we use the unemployment rate to measure the business cycle. The unemployment rate measures 11% on average. In terms of age we find the largest share between 30 and 45 years. The smallest share is found for the oldest age group, which is not surprising given that individuals start to draw a pension with 65-67, depending on their birth year. Finally, we use the age-adjusted death rate in order to proxy for average health within each state. Here we have an average of 503 deaths per 100,000 inhabitants.

Table 5.2: Summary Statistics, Panel of 16 states, 1999-2011 (2006 missing)

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Non-Infectious	70.40	5.83	57.86	87.69	192
Male Non-Infect.	71.19	6.33	60.86	92.28	192
Female Non-Infect.	71.35	7.25	55.42	90.18	192
Infectious	8.92	2.06	4.02	15.96	192
Male Infect.	8.82	2.23	3.9	16.61	192
Female Infect.	9.02	1.92	4.29	15.21	192
Respiratory	32.36	3.86	25.09	47.54	192
Male Resp.	30.21	4.2	22.45	46.15	192
Female Resp.	34.83	3.99	26.38	49.25	192
Members (in Thousands)	330.86	282.47	35.91	995	192
Share Female	45.31	3.86	30.49	50.97	192
Unemployment	11.14	4.5	3.8	20.5	192
Male Unemp.	11.19	4.44	3.7	21.1	192
Female Unemp.	11.06	4.75	3.9	23.3	192
Share Aged 30 minus	22.38	1.62	18.98	27.16	192
Share Aged 30-45	40.83	4.05	30.91	47.69	192
Share Aged 45-60	33.35	4.15	26.65	43.39	192
Share Aged 60 plus	3.44	0.94	1.22	5.84	192
Death Rate	503.03	42.16	399.3	609	192

This table shows the summary statistics. Sickness absence is divided in three groups, based on the cause. These groups consist of non-infectious diseases (group 1), infectious diseases (group 2) and diseases of the respiratory system (group 3). For each group we present the sickness spells started per 100 members. Moreover, we present overall sickness spells and spells divided by gender.

5.4.2 Empirical specification

Our baseline looks as follows:

$$S_{jt} = \alpha + \beta E_{jt} + \gamma D_{jt} + \epsilon_{jt}, \tag{5.5}$$

where S_{jt} stands for the natural logarithm of sickness absence in state j at time t. In terms of sickness absence we consider the number of sickness spells started. This is in accordance with our model above, since we have no clear predictions in terms of the length of sickness absence. The variable E_{jt} captures the economic condition in state j at time t. As is standard in this literature, we will use the unemployment rate as a proxy for the business cycle. Furthermore, we control for demographic characteristics D_{jt}. Expressing our hypothesis using the equation above, we expect a higher β (in absolute value) for infectious diseases than for non-infectious diseases.

We estimate different specifications and also add further variables. First, we control for the age-adjusted death rate as a proxy for average health in each state. Moreover, we also add specifications with year dummies. Infectious diseases such as the flu might appear in (yearly) cycles with different intensities, due to mutation of the virus. Since these cycles are unrelated to unemployment they would add unnecessary noise to our model. Third, there could be unobserved heterogeneity at state level. For instance general health, availability of doctors and other variables influencing health might differ from state to state. Moreover, standard variables that could influence individual absenteeism could differ by state, examples of such variables are martial status and how many children individuals have on average. Therefore, we also estimate a fixed effects model. These specifications are more demanding since most of the variation will be captured by the state and year dummies. Standard errors will be clustered at state level. We also estimated robust standard errors. However, they were smaller than the clustered standard errors displayed here.

Before we analyze the results we will briefly discuss potential endogeneity problems. Sickness absence could be due to chronic illnesses that might lead to a decrease in the unemployment rate due to individuals dropping out of the working force. However, such cases of sickness absence are not included in our data, since we only look at sickness absence of paying members of health insurance funds.

Another potential threat is reverse causality in the sense that sickness absence might lead to unemployment. However, this would create a positive link between the two variables, and thus it would only lead to an underestimation of the true effect. Moreover, this should not harm our identification since the mechanism should not differ between infectious and non-infectious diseases.

5.5 Results

5.5.1 Regression Results

Table 5.3 displays the results. In Panel A we present the results for non-infectious diseases. In the first column we present the results of an OLS regression. In the second column we include the death rate as a proxy for health. Columns 3 and 4 include time dummies. Finally, the last two columns show the results from the fixed effects regression.

The results show that older individuals exhibit higher sickness absence. Surprisingly in the regressions a higher share of females leads to less sickness absence. However, this might be related to the fact that we separate sickness absence by its cause.[61] Moreover, a higher death rate increases sickness absence. Finally, we find a small positive influence of unemployment on sickness absence of non-infectious diseases. This positive effect meets the predictions of our theoretical model, where sickness absence of non-infectious diseases increases during booms, due to less people staying at home when they are sick. However, this result is not significant in all specifications. An explanation for this non-significance can be found in Bell et al. (2012), who show that working more hours than desired has negative effects on health. This is also supported by the medical literature. Many studies in this literature find that sickness presenteeism leads to higher future sickness absence (see for instance Bergström, Bodin, Hagberg, Aronsson and Josephson 2009, Bergström, Bodin, Hagberg, Lindh, Aronsson and Josephson 2009, Hansen and Andersen 2009).

In Panel B we show the results for infectious diseases. Here the influence of unemployment is much larger and highly significant in all specifications. Since unemployment is countercyclical, previous empirical results would predict a negative relationship between unemployment and sickness absence. This is what we find in the results. Quantitatively we find that an increase of the unemployment rate by one percentage point will decrease sickness absence of infectious diseases by 2.5-4%. Moreover, there are almost no differences in sickness absence in terms of gender and age. The death rate has also little influence in most specifications, especially after controlling for unobserved heterogeneity at state level. This finding is not surprising since infectious diseases occur for all ages. While it is true that the immun system of elderly individuals might be more affected leading to longer sickness absence, we only measure the number of sickness spells started and not the length.

Finally, Panel C presents the results for respiratory diseases. Again the estimates for unemployment are quite small in most specifications. The share of females, the age group and the death rate also show no significant influence.

[61] For studies explaining higher female sickness absence see Laaksonen et al. (2008).

Table 5.3: Regression Results

Panel A: Non-Infectious Diseases (Group 1) as Dependent Variable

	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment	0.00744**	0.000355	0.00637*	0.00289	-0.00436	-0.00514
	0.00304	0.00330	0.00316	0.00417	0.00352	0.00373
Share Female	-0.0143***	-0.00956**	-0.00995**	-0.00784**	-0.0182***	-0.0187***
	0.00401	0.00390	0.00365	0.00322	0.00523	0.00554
Share Aged 30-45	-0.00787	-0.0138	0.00162	-0.00160	0.0330*	0.0307*
	0.00778	0.00893	0.0107	0.0124	0.0169	0.0158
Share Aged 45-60	-0.00527	-0.00328	0.00651	0.00634	0.0538***	0.0523***
	0.00752	0.00864	0.00687	0.00731	0.0159	0.0149
Share Aged 60 plus	0.0305	0.0361*	0.0191	0.0227	0.0494*	0.0494*
	0.0186	0.0195	0.0215	0.0227	0.0267	0.0266
Death Rate		0.00142***		0.000752		0.000422
		0.000379		0.000681		0.000560
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	No	No	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	No	Yes	Yes
Observations	192	192	192	192	192	192
R^2	0.247	0.328	0.602	0.619	0.768	0.769

Panel B: Infectious Diseases (Group 2) as Dependent Variable

	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment	-0.0258***	-0.0387***	-0.0246***	-0.0329***	-0.0384***	-0.0399***
	0.00649	0.00855	0.00735	0.0105	0.00751	0.00841
Share Female	0.00591	0.0145	0.00287	0.00789	0.0108	0.00984
	0.00749	0.00896	0.00982	0.0101	0.0123	0.0122
Share Aged 30-45	0.0292**	0.0184	0.0361*	0.0284	0.0410	0.0366
	0.0111	0.0120	0.0190	0.0186	0.0334	0.0304
Share Aged 45-60	0.0158	0.0194	0.0248*	0.0244*	0.0351	0.0323
	0.0118	0.0122	0.0129	0.0136	0.0326	0.0314
Share Aged 60 plus	0.139***	0.149***	0.0849**	0.0935**	-0.00792	-0.00790
	0.0340	0.0358	0.0365	0.0322	0.0466	0.0464
Death Rate		0.00260***		0.00180		0.000807
		0.000829		0.00119		0.00112
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	No	No	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	No	Yes	Yes
Observations	192	192	192	192	192	192
R^2	0.453	0.491	0.758	0.771	0.859	0.860

Panel C: Respiratory Diseases (Group 3) as Dependent Variable

	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment	-0.00623	-0.00932	-0.00990**	-0.00854	0.00134	0.00102
	0.00485	0.00559	0.00366	0.00649	0.00542	0.00503
Share Female	-0.00785	-0.00580	0.00274	0.00192	-0.00491	-0.00511
	0.00583	0.00567	0.00546	0.00480	0.00805	0.00862
Share Aged 30-45	-0.00480	-0.00737	0.0131	0.0144	0.0109	0.00993
	0.0137	0.0146	0.0142	0.0161	0.0187	0.0163
Share Aged 45-60	-0.00302	-0.00215	0.0131	0.0132*	0.0260	0.0254
	0.0126	0.0131	0.00779	0.00747	0.0193	0.0175
Share Aged 60 plus	0.00882	0.0112	0.000412	-0.000980	0.0406	0.0406
	0.0210	0.0225	0.0312	0.0321	0.0307	0.0307
Death Rate		0.000621		-0.000293		0.000176
		0.000582		0.000970		0.000708
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	No	No	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	No	Yes	Yes
Observations	192	192	192	192	192	192
R^2	0.110	0.119	0.595	0.597	0.745	0.745

The regressions are weighed by the average number of members in each state. The second row presents robust standard errors. The stars represent significance at the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.4: Regression Results for Infectious Diseases Divided by Gender

	Males			Females		
	(1)	(2)	(3)	(1)	(2)	(3)
Unemployment	-0.0416***	-0.0345**	-0.0440***	-0.0339***	-0.0294***	-0.0331***
	0.0103	0.0126	0.0111	0.00612	0.00762	0.00625
Share Female	0.0138	0.00538	0.00526	0.0146	0.00972	0.0137
	0.00884	0.00988	0.0124	0.00926	0.0104	0.0121
Share Aged 30-45	0.0301**	0.0364	0.0531	0.00631	0.0197	0.0217
	0.0137	0.0213	0.0361	0.0117	0.0163	0.0251
Share Aged 45-60	0.0193	0.0219	0.0521	0.0184	0.0258**	0.0146
	0.0141	0.0166	0.0366	0.0110	0.0109	0.0259
Share Aged 60 plus	0.192***	0.132***	-0.000198	0.109***	0.0553*	-0.00357
	0.0362	0.0374	0.0510	0.0349	0.0314	0.0436
Death Rate	0.00241**	0.00142	0.000498	0.00266***	0.00201*	0.000800
	0.000955	0.00139	0.00113	0.000699	0.000990	0.00119
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	No	Yes	Yes	No	Yes	Yes
Fixed Effects	No	No	Yes	No	No	Yes
Observations	192	192	192	192	192	192
R^2	0.512	0.755	0.841	0.471	0.784	0.874

The regressions are weighted by the average number of members in each state. The second row presents robust standard errors. The stars represent significance at the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This finding is more surprising, and we will address it in the discussion below.

5.5.2 Robustness Checks

Before we discuss the results we perform some robustness checks. First, we divide our results for infectious diseases by gender in order to see whether the mechanism appears for females and males. For these regressions we use data on sickness absence and unemployment divided by gender. The results, shown in Table 5.4, suggest that the mechanism appears for both males and females. Moreover, the effect seems more pronounced for men.

In order to exclude spurious results we first difference our data. In the case of serial correlation, first differencing of the data will lead to a higher efficiency. Table 5.5 displays the results of the first differenced data.

Due to the fact that the year 2006 is missing for our dependent variables, the number of observations drop. The point estimates for unemployment are again significant and negative. Quantitatively the effect is a little larger at first, however, once we include year dummies the results are quite similar to our previous estimates with a coefficient of around -4%.

Summing up, the data seems to confirm the hypothesis we derived from the theoretical model. Only for infectious diseases we find a procyclical relationship as predicted by our theoretical model.

Table 5.5: Regression Results for Infectious Diseases (Group 2) After First Differencing the Data

	(1)	(2)	(3)	(4)
Unemployment	-0.0765***	-0.0787***	-0.0402**	-0.0390**
	0.0196	0.0200	0.0165	0.0160
Share Female	0.0180	0.0185	0.0112	0.0108
	0.0202	0.0203	0.0165	0.0165
Share Aged 30-45	0.0949*	0.0936*	0.0744*	0.0693*
	0.0495	0.0493	0.0375	0.0369
Share Aged 45-60	-0.0302	-0.0354	-0.00563	-0.00829
	0.0763	0.0745	0.0459	0.0457
Share Aged 60 plus	0.202***	0.204**	-0.0744	-0.0751
	0.0678	0.0694	0.0867	0.0866
Death Rate		0.00145**		0.00110
		0.000624		0.00106
[1em] Constant	Yes	Yes	Yes	Yes
Year Dummies	No	No	Yes	Yes
Observations	160	160	160	160
R^2	0.275	0.281	0.844	0.845

The regressions are weighted by the average number of members in each state. The second row presents robust standard errors. The stars represent significance at the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.6 Discussion

Our results suggest sickness absence due to infectious diseases is procyclical, while non-infectious diseases show hardly any relation to the business cycle. In terms of which explanation is the driving force behind the procyclical nature of sickness absence we cannot reject the labor force composition explanations. Since our data includes both employed and unemployed individuals and we cannot distinguish between them, we cannot test whether the sickness absence would be even stronger when excluding unemployed individuals. Finally, we do not find any support for the moral hazard explanation in our data, since moral hazard can not explain why the relationship is only found for infectious diseases. Rather, one would expect to find a more pronounced relationship to the business cycle independent of the cause of sickness if moral hazard was the driving force.

While our findings suggest that infectious diseases are procyclical, it is surprising that the relationship does not arise for diseases of the respiratory system. For diseases like common colds and other diseases of the respiratory system there seems to be no relation with the business cycle. There are several possible explanations for this finding. Firstly, diseases of the respiratory system consist of both infectious and non-infectious diseases. Therefore, this finding could be driven by non-infectious diseases contained in this category. However,

as seen at the end of Section 5.4.1 most sickness cases in this category are due to infectious diseases. A more likely explanation is that the model parameters are such that sickness absence is not procyclical (and thus Proposition 1 will not hold). For instance, one could argue that the fraction q of individuals going to work while sick is comparatively high for some diseases in this category, such as a common cold. This seems likely especially when considering that most cases for sickness absence are only registered if they last longer than three days. If this parameter is too high, there will be many infections, however only quite few individuals stay at home causing only very small or no cyclical patterns of sickness absence. Another reason for this finding could be that the exogenous sickness probability p_{ex} is quite high for diseases of the respiratory system. The flu usually appears in cycles with very high prevalence rates every winter. For instance, Turkington and Ashby (2007) estimate that every year up to 50 million people will be infected in the US. During times of very high prevalence infections may be very high and this might cause the observation that we observe no relationship with the business cycle. One would need higher frequency data in order to investigate this point further.

In this Chapter we analyzed the question why sickness absence increases during periods of economic growth. The literature so far suggests two possible explanations. The first is driven by the composition of the labor force, with individuals with a lower health status who only work in times of economic booms. A second explanation is given by moral hazard, since employees might have a lower fear of job loss in more expansive periods.

We provide a third explanation caused by a heavier workload, resulting in more presenteeism and infections during a period of favorable economic conditions. We first present the theoretical underpinnings of these explanations and then test the explanation using sickness absence data. Since our data includes categories showing the cause of the sickness, we are able to identify what cause mainly drives the negative relationship between sickness absence and unemployment. We find that infectious diseases show the most pronounced association.

In terms of policy recommendation, the results suggest that there are hidden costs of incentive wages and other forms of motivation, since they might lead to more presenteeism, especially when the economy is booming and sales are high. As shown in this paper, presenteeism will lead to infection of co-workers, causing overall higher absence rates during economic booms. In order to prevent this mechanism it seems promising to encourage employees to go on sickness absence, in particular if the distress is caused by infectious diseases, to avoid costly infections and possible pandemics.

CHAPTER

6

Conclusion

This dissertation analyzed the interactions between infectious diseases and the economy. The different Chapters theoretically and empirically examined different channels of interaction. Within each Chapter we initially provided a theoretical foundation which built upon extensions and adaptations of existing models and partly new models. In a second step we presented the empirical analyses and the corresponding results using modern empirical methods. We analyzed HIV/AIDS, the Spanish Flu and infectious diseases at the workplace. HIV/AIDS and the Spanish Flu are special since they mainly affect prime-age individuals, and thus have considerable effects on labor supply. In the first half of this dissertation we investigated how infectious diseases affect the economy. In the second half the reverse relationship was analyzed, namely how the business cycle affects the spread of diseases. In this last Chapter, we briefly summarize the results and investigate paths for further research.

The second Chapter analyzed how HIV prevalence affects the Gross Domestic Product (GDP) in the twelve most heavily affected countries, i.e. the countries which have ever exhibited a HIV prevalence rate of 10% or more. Here the analysis was divided into a theoretical approach following a Solow growth model and an empirical approach using data and results from previous empirical studies on how different determinants affect the growth path of a country. The twelve countries were analyzed by means of synthetic control groups, a new method developed by Abadie and Gardeazabal (2003). For every country under consideration a synthetic control group was formed, consisting

of a weighted average of donor countries not affected by HIV (HIV prevalence rate smaller than 1%). The results in the different countries were very heterogeneous. While some countries showed no effect at all, other countries exhibited a reduction in GDP between 25 and 77%, compared to the scenario without HIV.

Since we revealed very different effects in Chapter 2, we tried to restrict the possible channels in Chapter 3 by analyzing demographic effects of HIV. More specifically, in this Chapter we estimated how HIV affects life expectancy, and death and birth rates. In order to find better control groups the synthetic control group method was extended to several variables. As expected there were very clear-cut effects on life expectancy and the death rate. Average life expectancy decreased by almost 15 years, while the death rate increased by seven deaths on average (per 1,000 inhabitants) due to HIV. In terms of the birth rate the effects were once more quite heterogeneous. On average we estimated a very small and statistically insignificant effect, while in Zambia a positive effect was found.

Zambia was also the country where we revealed the largest effects in terms of GDP. Combining the findings of Chapters 2 and 3 we found that the channel working through fertility seems very important. Most countries in the world experienced a drop in fertility over the last decade. Chen (2010) suggests that this is due to increased life expectancy and a more concentrated investment in fewer children. This pattern seems to be missing for some African countries affected by HIV. However, other countries of high HIV prevalence such as Swaziland, experienced drop in terms of fertility. In general we found that the effects of HIV on GDP were intensified by a non-decline in the fertility rate.

Summing up, the results call for more case studies, in order to analyze in more depth the effects in the single countries. This is needed to better understand the mechanisms and the policies for preventing severe drops in GDP. Some case studies can already be found in the literature. See, for instance, Arndt (2006) for a case study in Mozambique, MacFarlan and Sgherri (2006) for Botswana and Robalino et al. (2002) for Kenya. However, more research is needed to obtain a clearer picture.

Due to the effect heterogeneity between countries observed in the last Chapters, we provided an in depth analysis of one country in Chapter 4. One reason for the effect heterogeneity is that different countries took different measures to prevent the spread of the disease. Therefore, in this Chapter we analyzed the economic consequences of the Spanish Flu – a disease that occurred very rapidly and thus prevention was very limited. The disease spread between 1918 and 1920. Due to the time period it is difficult to separate effects of the Spanish Flu from the consequences of World War I. Therefore the analysis focused

on a country which was neutral during the War: Sweden. Other reasons for analyzing the effects on Sweden were the good data availability and vast differences in terms of prevalence and mortality between Swedish counties. These differences were used to measure the effects of the Spanish Flu. The results showed that capital income decreased more in counties with more flu-related deaths, while poverty increased. However, there was robust evidence that the influenza had no discernible effect on earnings. This finding was surprising since it goes against most previous empirical studies as well as theoretical predictions. We suggested how these findings relate to the growth literature and proposed that an externality on capital could explain the findings. However, more theoretical research is needed here also.

In Chapter 5 the causal channel was reversed and we analyzed how the business cycle affects the spread of diseases. The procyclical nature of sickness absence has been documented by many scholars in the literature. So far, the explanations have been based on labor force composition and moral hazard. We proposed and tested a third mechanism caused by presenteeism (i.e. working while sick) and infections. We suggested that the workload is higher during an economic boom and thus employees go to work even when they do not feel well. In a theoretical model focusing on infectious diseases, we showed that this provokes infections of co-workers leading to overall higher sickness absence during economic upturns. Using county-level aggregated data from 112 German public sickness funds (out of 145 in total) we confirmed this hypothesis by finding that infectious diseases show the largest procyclical pattern. For this explanation further empirical studies are needed as well. Firstly, higher frequency data (e.g. monthly data) would allow to study the dynamics of infections more in detail. Moreover, unemployment is probably not the best indicator here. Data on individual working hours would provide an even clearer picture about the mechanism. Finally, studies in other countries are needed in order to investigate the external validity of our findings.

Bibliography

Abadie, A., Diamond, A. and Hainmueller, J.: 2010, Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program, *Journal of the American Statistical Association* **105**(490), 493–505.

Abadie, A. and Gardeazabal, J.: 2003, The economic costs of conflict: A case study of the Basque Country, *The American Economic Review* **93**(1), 113–132.

Acemoglu, D.: 2003, Labor-and capital-augmenting technical change, *Journal of the European Economic Association* **1**(1), 1–37.

Acemoglu, D., Johnson, S. and Robinson, A.: 2001, The colonial origins of comparative development: an empirical investigation, *The American Economic Review* **91**(5), 1369–1401.

Acemoglu, D., Johnson, S. and Robinson, J.: 2003, Disease and development in historical perspective, *Journal of the European Economic Association* **1**(2-3), 397–405.

Alan Heston, A., Summers, R. and Aten, B.: 2006, Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Albouy, D.: 2008, The colonial origins of comparative development: An investigation of the settler mortality data, NBER Working Papers.

Almond, D.: 2006, Is the 1918 influenza pandemic over? Long-term effects

of in utero influenza exposure in the post-1940 US population, *Journal of Political Economy* **114**(4), 672–712.

Almond, D. and Mazumder, B.: 2005, The 1918 influenza pandemic and subsequent health outcomes: An analysis of SIPP data, *The American Economic Review* **95**(2), 258–262.

Åman, M.: 1990, *Spanska sjukan: Den svenska epidemin 1918-1920 och dess internationella bakgrund*, Ubsaliensis Academiae.

American Medical Association: 2012, Fact sheet: The differences between ICD-9 and ICD-10, Last Updated September 25.

Anderson, J.: 1979, A theoretical foundation for the gravity equation, *The American Economic Review* **69**(1), 106–116.

Angrist, J. and Pischke, J.: 2008, *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press.

Arai, M. and Thoursie, P.: 2005, Incentives and selection in cyclical absenteeism, *Labour Economics* **12**(2), 269–280.

Arellano, M.: 2003, *Panel Data Econometrics*, Oxford University Press.

Arellano, M. and Bond, S.: 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The Review of Economic Studies* **58**(2), 277–297.

Armstrong, H. and Read, R.: 1998, Trade and growth in small states: the impact of global trade liberalisation, *The World Economy* **21**(4), 563–585.

Arndt, C.: 2006, HIV/AIDS, human capital, and economic growth prospects for Mozambique, *Journal of Policy Modeling* **28**(5), 477–489.

Aronsson, G. and Gustafsson, K.: 2005, Sickness presenteeism: Prevalence, attendance-pressure factors, and an outline of a model for research, *Journal of Occupational and Environmental Medicine* **47**(9), 958–966.

Aronsson, G., Gustafsson, K. and Dallner, M.: 2000, Sick but yet at work. An empirical study of sickness presenteeism, *Journal of Epidemiology and Community Health* **54**(7), 502–509.

Arrehag, L., De Vylder, S., Durevall, D. and Sjöblom, M.: 2006, *The impact of HIV/AIDS on livelihoods, poverty and the economy of Malawi*, Sida.

Askildsen, J., Bratberg, E. and Nilsen, Ø.: 2005, Unemployment, labor force composition and sickness absence: A panel data study, *Health Economics* **14**(11), 1087–1101.

Audas, R. and Goddard, J.: 2001, Absenteeism, seasonality, and the business cycle, *Journal of Economics and Business* **53**(4), 405–419.

Banerjee, A., Amsden, A., Bates, R., Bhagwati, J., Deaton, A. and Stern, N.: 2007, *Making Aid Work*, Vol. 1, The MIT Press.

Barmby, T. and Larguem, M.: 2009, Coughs and sneezes spread diseases: an empirical study of absenteeism and infectious illness, *Journal of Health Economics* **28**(5), 1012–1017.

Barro, R.: 1991, Economic growth in a cross section of countries, *The Quarterly Journal of Economics* **106**(2), 407.

Barro, R.: 2001, Human capital and growth, *The American Economic Review* **91**(2), 12–17.

Barry, J.: 2005, *The great influenza: The epic story of the deadliest plague in history*, Penguin Group USA.

Becker, G., Murphy, K. and Tamura, R.: 1990, Economic growth, human capital and population growth, *Journal of Political Economy* **98**(5), S12–S137.

Bell, C., Devarajan, S. and Gersbach, H.: 2006, The long-run economic costs of AIDS: A model with an application to South Africa, *The World Bank Economic Review* **20**(1), 55.

Bell, D., Otterbach, S. and Sousa-Poza, A.: 2012, Work hours constraints and health, *Annales d'Economie et de Statistique* **105–106**, 35–54.

Bergh, A.: 2007, *Den kapitalistiska välfärdsstaten*, Norstedts akademiska förlag.

Bergström, G., Bodin, L., Hagberg, J., Aronsson, G. and Josephson, M.: 2009, Sickness presenteeism today, sickness absenteeism tomorrow? A prospective study on sickness presenteeism and future sickness absenteeism, *Journal of Occupational and Environmental Medicine* **51**(6), 629–638.

Bergström, G., Bodin, L., Hagberg, J., Lindh, T., Aronsson, G. and Josephson, M.: 2009, Does sickness presenteeism have an impact on future general health?, *International Archives of Occupational and Environmental Health* **82**(10), 1179–1190.

Bertocchi, G. and Canova, F.: 2002, Did colonization matter for growth? An empirical exploration into the historical causes of Africa's underdevelopment, *European Economic Review* **46**(10), 1851–1871.

Bertrand, M., Duflo, E. and Mullainathan, S.: 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics* **119**(1), 249–275.

Bloom, D. and Mahal, A.: 1997a, AIDS, flu, and the Black Death: Impacts on economic growth and well-being, *The Economics of HIV and AIDS: The Case of South and South East Asia*, Delhi: Oxford University Press, pp. 22–52.

Bloom, D. and Mahal, A.: 1997b, Does the AIDS epidemic threaten economic growth?, *Journal of Econometrics* **77**(1), 105–124.

Blundell, R. and Bond, S.: 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* **87**(1), 115–143.

Bongaarts, J., Pelletier, F. and Gerland, P.: 2010, How many more AIDS deaths?, *The Lancet* **375**(9709), 103–104.

Bos, K., Schuenemann, V., Golding, G., Burbano, H., Waglechner, N., Coombes, B., McPhee, J., DeWitte, S., Meyer, M., Schmedes, S. et al.: 2011, A draft genome of yersinia pestis from victims of the black death, *Nature* **478**(7370), 506–510.

Boucekkine, R., Croix, D. and Peeters, D.: 2007, Early literacy achievements, population density, and the transition to modern growth, *Journal of the European Economic Association* **5**(1), 183–226.

Boucekkine, R., De la Croix, D. and Licandro, O.: 2002, Vintage human capital, demographic trends, and endogenous growth, *Journal of Economic Theory* **104**(2), 340–375.

Boucekkine, R. and Laffargue, J.: 2010, On the distributional consequences of epidemics, *Journal of Economic Dynamics and Control* **34**(2), 231–245.

Boucekkine, R. and Ruiz Tamarit, R.: 2005, Imbalance effects in the Lucas model: An analytical exploration, *Topics in Macroeconomics* **4**(1).

Brainerd, E. and Siegler, M.: 2003, *The economic effects of the 1918 influenza epidemic*, number 3791, Centre for Economic Policy Research.

Branke, J., Deb, K., Miettinen, K. and Slowinski, R.: 2008, *Multiobjective optimization: Interactive and evolutionary approaches*, Vol. 5252, Springer.

Bratberg, E. and Monstad, K.: 2012, Worried sick? Worker responses to organizational turmoil, mimeo.

Brown, R.: 2010, The 1918 US influenza pandemic as a natural experiment, revisited, mimeo.

Buettner, T., Sawyer, C. and Zlotnik, H.: 2003, Assessing the adequacy of AIDS modelling in the estimation of population, *scientific meeting on empirical evidence for the demographic and socio-economic impact of AIDS, hosted by HEARD, Durban, South Africa*, pp. 26–28.

Byrne, J.: 2006, *Daily life during the Black Death*, Greenwood Publishing Group.

Cahu, P. and Fall, F.: 2011, Accounting for the effects of AIDS on growth in Sub-Saharan Africa, Université Paris1 Panthéon-Sorbonne (Post-Print and Working Papers).

Capasso, V.: 2008, *Mathematical structures of epidemic systems*, Vol. 97, Springer.

Chen, H.: 2010, Life expectancy, fertility, and educational investment, *Journal of Population Economics* **23**(1), 37–56.

Cheung, P.: 2008, Undata database: Capital formation (constant prices), UN (United Nations).

Chowell, G., Bettencourt, L., Johnson, N., Alonso, W. and Viboud, C.: 2008, The 1918–1919 influenza pandemic in England and Wales: Spatial patterns in transmissibility and mortality impact, *Proceedings of the Royal Society B: Biological Sciences* **275**(1634), 501.

Ciccone, A. and Jarocinski, M.: 2010, Determinants of economic growth: will data tell?, *American Economic Journal: Macroeconomics* **2**(4), 222–246.

Cingano, F. and Schivardi, F.: 2004, Identifying the sources of local productivity growth, *Journal of the European Economic Association* **2**(4), 720–744.

Cockburn, A. et al.: 1963, *The evolution and eradication of Infectious Diseases*, Cambridge University Press.

Collier, P. and Gunning, J.: 1999a, Explaining African economic performance, *Journal of Economic Literature* **37**(1), 64–111.

Collier, P. and Gunning, J.: 1999b, Why has Africa grown slowly?, *The Journal of Economic Perspectives* **13**(3), 3–22.

Coneus, K. and Mühlenweg, A.: 2011, Orphans at risk in Sub-Saharan Africa: Evidence on educational and health outcomes, ZEW Discussion Papers.

Cutler, D., Deaton, A. and Lleras-Muney, A.: 2006, The determinants of mortality, National Bureau of Economic Research.

Davies, R., Jones, P. and Nuñez, I.: 2009, The impact of the business cycle on occupational injuries in the UK, *Social Science & Medicine* **69**(2), 178–182.

Demerouti, E., Le Blanc, P., Bakker, A., Schaufeli, W. and Hox, J.: 2009, Present but sick: a three-wave study on job demands, presenteeism and burnout, *Career Development International* **14**(1), 50–68.

Dixit, A.: 1973, The optimum factory town, *The Bell Journal of Economics and Management Science* pp. 637–651.

Donald, S. and Lang, K.: 2007, Inference with difference-in-differences and other panel data, *The Review of Economics and Statistics* **89**(2), 221–233.

Dosi, G. and Nelson, R.: 2010, Technical change and industrial dynamics as evolutionary processes, Handbook of Innovation.

Dreher, A.: 2006, Does globalization affect growth? Evidence from a new index of globalization, *Applied Economics* **38**(10), 1091–1110.

Durevall, D. and Lindskog, A.: 2009, Economic inequality and HIV in Malawi, *Rapport Nr.: Working Papers in Economics 425* .

Durevall, D. and Lindskog, A.: 2011, Uncovering the impact of the HIV epidemic on fertility in Sub-Saharan Africa: the case of Malawi, *Journal of Population Economics* **24**(2), 629–655.

Durlauf, S., Kourtellos, A. and Tan, C.: 2008, Are any growth theories robust?, *The Economic Journal* **118**(527), 329–346.

Dyrstad, J. and Ose, S.: 2002, Non-linear unemployment effects in sickness absence: Discipline or composition effects?, *Working Paper Series* (25).

Elmér, Å.: 1960, *Folkpensioneringen i Sverige, med särskild hänsyn till ålderspensioneringen: Old age pensions in Sweden*, Vol. 17, CWK Gleerup, Lund.

Englebert, P.: 2000, Solving the mystery of the Africa dummy, *World development* **28**(10), 1821–1835.

Englebert, P.: 2002, *State Legitimacy and Development in Africa*, Lynne Rienner Pub.

Fabiani, M., Nattabi, B., Ayella, E., Ogwang, M. and Declich, S.: 2006, Differences in fertility by HIV serostatus and adjusted HIV prevalence data from an antenatal clinic in northern Uganda, *Tropical Medicine & International Health* **11**(2), 182–187.

Fahr, R. and Frick, B.: 2007, On the inverse relationship between unemployment and absenteeism: Evidence from natural experiments and worker heterogeneity, IZA Discussion Paper No. 3171.

Fearon, J.: 2003, Ethnic and cultural diversity by country, *Journal of Economic Growth* **8**(2), 195–222.

Fink, G. and Linnemayr, S.: 2008, HIV, education and fertility: long term evidence from Sub-Saharan Africa, Working paper, Harvard School of Public Health.

Fogel, R.: 1994, Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy1, *The American Economic Review* **84**(3), 369–395.

Fogel, R.: 2004, *The escape from hunger and premature death, 1700-2100: Europe, America, and the Third World*, Vol. 38, Cambridge University Press.

Fortson, J.: 2009, HIV/AIDS and fertility, *American Economic Journal: Applied Economics* **1**(3), 170–194.

Fortson, J.: 2011, Mortality risk and human capital investment: The Impact of HIV/AIDS in Sub-Saharan Africa, *The Review of Economics and Statistics* **93**(1), 1–15.

Francis, P.: 2004, Optimal tax/subsidy combinations for the flu season, *Journal of Economic Dynamics and Control* **28**(10), 2037–2054.

Gaffeo, E.: 2003, The economics of HIV/AIDS: a survey, *Development Policy Review* **21**(1), 27–49.

Gallup, J., Mellinger, A. and Sachs, J.: 2001, Geography datasets, Center for International Development, Harvard University.

Garrett, T.: 2008, Pandemic economics: The 1918 influenza and its modern-day implications, *Federal Reserve Bank of St. Louis Review* **90**(2), 75–93.

Garrett, T.: 2009, War and pestilence as labor market shocks: US manufacturing wage growth 1914–1919, *Economic Inquiry* **47**(4), 711–725.

German Federal Ministry of Health: 2012, Mitglieder und Versicherte, http://bundesgesundheitsministerium.de/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung.html, last accessed November 2, 2012.

Gersovitz, M. and Hammer, J. S.: 2004, The economical control of infectious diseases, *The Economic Journal* **114**(492), 1–27.

Goenka, A. and Liu, L.: 2012, Infectious diseases and endogenous fluctuations, *Economic Theory* **50**(1), 125.

Goetzel, R., Long, S., Ozminkowski, R., Hawkins, K., Wang, S. and Lynch, W.: 2004, Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting us employers, *Journal of Occupational and Environmental Medicine* **46**(4), 398–412.

Hansen, C. and Andersen, J.: 2008, Going ill to work–what personal circumstances, attitudes and work-related factors are associated with sickness presenteeism?, *Social Science & Medicine* **67**(6), 956–964.

Hansen, C. and Andersen, J.: 2009, Sick at work–a risk factor for long-term sickness absence at a later date?, *Journal of Epidemiology and Community Health* **63**(5), 397–402.

Haub, C.: 2010, *World Population Data Sheet of the Population Reference Bureau, Inc*, Population Reference Bureau.

Hausman, J. and Kuersteiner, G.: 2008, Difference in difference meets generalized least squares: Higher order properties of hypotheses tests, *Journal of Econometrics* **144**(2), 371–391.

Helsley, R. and Strange, W.: 1991, Agglomeration economies and urban capital markets, *Journal of Urban Economics* **29**(1), 96–112.

Hinrichs, P.: 2011, The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities, The Review of Economics and Statistics.

Hopkins, D.: 2002, *The greatest killer: smallpox in history*, University of Chicago Press.

Hultkvist, G.: 1940, Några anmärkningar till vår nya dödsorsaksstatistik, *Allmänna Svenska Läkartidningen* **17**, 58–60.

Hyrenius, H.: 1914, Den svenska officiella dödsorsaksstatistiken, *Social-medicinsk tidskrift* **11**, 1179–1188.

Ichino, A. and Riphahn, R.: 2005, The effect of employment protection on worker effort: Absenteeism during and after probation, *Journal of the European Economic Association* **3**(1), 120–143.

Influensabyrån: 1919, *Koncept. Registratur 1918–1919*, Stockholm: Kungl. Medicinalstyrelsen.

Influensakommittén: 1924, Influensakommitténs rapport, *Svenska Läkartidningen* **10**, 1–32.

Jerven, M.: 2009, The quest for the African dummy: explaining African post-colonial economic performance revisited, Journal of International Development.

Johansson, P. and Palme, M.: 1996, Do economic incentives affect work absence? Empirical evidence using Swedish micro data, *Journal of Public Economics* **59**(2), 195–218.

Johansson, P. and Palme, M.: 2005, Moral hazard and sickness insurance, *Journal of Public Economics* **89**(9), 1879–1890.

Johnson, N. and Mueller, J.: 2002, Updating the accounts: Global mortality of the 1918-1920 "Spanish" influenza pandemic, *Bulletin of the History of Medicine* **76**(1), 105–115.

Jörberg, L. and Krantz, O.: 1978, *Ekonomisk och social politik i Sverige, 1850-1939*, Ekonomisk-historiska institutionen, Lunds universitet.

Jorner, U.: 2008, *Summa summarum – SCB:s första 150 år*, Statistiska Centralbyrån (SCB)-Tryck, Örebro.

Juhn, C., Kalemli-Ozcan, S. and Turan, B.: 2008, HIV and fertility in Africa: First evidence from population based surveys, *Technical report*, National Bureau of Economic Research.

Kalemli-Ozcan, S.: 2006, AIDS, reversal of the demographic transition and economic development: Evidence from Africa, National Bureau of Economic Research.

Kalemli-Ozcan, S.: 2012, AIDS, reversal of the demographic transition and economic development: evidence from Africa, *Journal of Population Economics* **25**(3), 871–897.

Karlsson, M., Nilsson, T. and Pichler, S.: 2012, What doesn't kill you makes you stronger?, Darmstadt Discussion Papers in Economics Nr. 211.

Karlsson, M. and Pichler, S.: 2012a, Demographic consequences of HIV, Unpublished Manuscript, Darmstadt University of Technology.

Karlsson, M. and Pichler, S.: 2012b, The economic impact of the HIV pandemic, Unpublished Manuscript, Darmstadt University of Technology.

Kermack, W. and McKendrick, A.: 1927, A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London. Series A* **115**(772), 700–721.

Kiefer, N.: 1980, Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance, *Journal of Econometrics* **14**(2), 195–202.

Krantz, O. and Schön, L.: 2007, *Swedish historical national accounts 1800-2000*, Almqvist & Wiksell International.

Kulikowski, M.: 2007, Justinian's flea: Plague, empire, and the birth of Europe, *History: Reviews of New Books* **35**(4), 148–148.

La Porta, R., Lopez-de Silanes, F., Shleifer, A. and Vishny, R.: 1999, The quality of government, *Journal of Law, Economics, and Organization* **15**(1), 222.

Laaksonen, M., Martikainen, P., Rahkonen, O. and Lahelma, E.: 2008, Explanations for gender differences in sickness absence: evidence from middle-aged municipal employees from finland, *Occupational and Environmental Medicine* **65**(5), 325–330.

Laxminarayan, R. and Brown, G.: 2001, Economics of antibiotic resistance: A theory of optimal use, *Journal of Environmental Economics and Management* **42**(2), 183–206.

Lechner, M.: 2010, The estimation of causal effects by difference-in-difference methods.

Lederberg, J.: 1997, Infectious disease as an evolutionary paradigm., *Emerging Infectious Diseases* **3**(4), 417.

Leigh, J.: 1985, The effects of unemployment and the business cycle on absenteeism, *Journal of Economics and Business* **37**(2), 159–170.

Lindbeck, A., Palme, M. and Persson, M.: 2006, Job security and work absence: Evidence from a natural experiment.

Little, T., Garofalo, C. and Williams, P.: 1918, The absence of the bacillus influenzæ in the exudate from the upper air-passages in the present epidemic., *The Lancet* **192**(4950), 34–34.

Liu, G., Guo, J. and Smith, S.: 2004, Economic costs to business of the HIV/AIDS epidemic, *Pharmacoeconomics* **22**(18), 1181–1194.

Lucas, R.: 1988, On the mechanics of economic development, *Journal of Monetary Economics* **22**(1), 3–42.

Lutz, W., Goujon, A., Samir, K. and Sanderson, W.: 2007, Reconstruction of populations by age, sex and level of educational attainment for 120 countries for 1970-2000.

Maccini, S. and Yang, D.: 2009, Under the weather: Health, schooling, and economic consequences of early-life rainfall, *American Economic Review* **99**(3), 1006–1026.

MacFarlan, M. and Sgherri, S.: 2006, The macroeconomic impact of HIV/AIDS in Botswana.

Maclaine, D.: 2006, *HIV/AIDS and Economic Growth: An Econometric Analysis with Particular Consideration of the Role of Education Capital Accumulation*, PhD thesis, thesis, Brisbane: School of Economics, University of Queensland, Australia, Australian Centre for Economic Research on Health.

Magnusson, L.: 1996, *Sveriges ekonomiska historia*, Rabén Prisma.

Mamelund, S.: 2006, A socially neutral disease? Individual social class, household wealth and mortality from Spanish influenza in two socially contrasting parishes in Kristiania 1918-19, *Social Science & Medicine* **62**(4), 923–940.

Marshall, M.: 2010, Major episodes of political violence 1946-2008.

Marshall, M., Jaggers, K. and Gurr, T.: 2009, Polity IV project: Political regime characteristics and transitions, 1800-2009.

McDonald, S. and Roberts, J.: 2006, AIDS and economic growth: A human capital approach, *Journal of Development Economics* **80**(1), 228–250.

McNeill, W.: 1977, *Plagues and peoples*, Anchor.

McSweeny, K., Colman, A., Fancourt, N., Parnell, M., Stantiall, S., Rice, G., Baker, M. and Wilson, N.: 2007, Was rurality protective in the 1918

influenza pandemic in New Zealand?, *Journal of the New Zealand Medical Association* **120**(1256).

Medizinischer Dienst der Krankenversicherung (MDK): 2012, Medizinischer Dienst der Krankenversicherung (MDK), www.mdk.de.

Millard, S., Scott, A. and Sensier, M.: 1997, The labour market over the business cycle: can theory fit the facts?, *Oxford Review of Economic Policy* **13**(3), 70–92.

Miller, D., Page, M., Stevens, A. and Filipski, M.: 2009, Why are recessions good for your health?, *The American Economic Review* pp. 122–127.

Montalvo, J.: 2011, Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections, *Review of Economics and Statistics* **93**(4), 1146–1154.

Morens, D. and Fauci, A.: 2007, The 1918 influenza pandemic: Insights for the 21st century, *Journal of Infectious Diseases* **195**(7), 1018.

Mundial, B.: 2011, *World development indicators*, The Wold Bank, IEC information Center, Development data group.

Murray, C., Salomon, J. and Mathers, C.: 2002, A critical examination of summary measures of population health, *in* C. Murray, J. Salomon, C. Mathers and A. Lopez (eds), *Summary Measures of Population Health*, Geneva: WHO Press, pp. 13–40.

Ndulu, B., O'Connell, S. and Azam, J.: 2008, *The Political Economy of Economic Growth in Africa, 1960-2000*, Vol. 2, Cambridge University Press.

Nelson, R.: 2010, Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic, *Health economics* **19**(10), 1181–1192.

Nordberg, M. and Røed, K.: 2009, Economic incentives, business cycles, and long-term sickness absence, *Industrial Relations: A Journal of Economy and Society* **48**(2), 203–230.

OECD: 2005, *Trends in International Migration: Annual Report 2004*, Paris: OECD.

Oster, E.: 2007, Routes of infection: Exports and HIV incidence in Sub-Saharan Africa.

Oster, E.: 2010, Estimating HIV prevalence and incidence in Africa from mortality data, *The BE Journal of Economic Analysis & Policy* **10**(1), 80.

Osterholm, M.: 2005, Preparing for the next pandemic, *New England Journal of Medicine* **352**(18), 1839–1842.

Patterson, K. and Pyle, G.: 1991, The geography and mortality of the 1918 influenza pandemic., *Bulletin of the History of Medicine* **65**(1), 4.

Petrén, E.: 1918a, Om influensans epidemiologi och profylax, *Allmänna Svenska Läkartidningen* **15**, 1630–1643.

Petrén, K.: 1918b, Till influensans, särskilt influensapneumoniernas nosografi och behandling, *Allmänna Svenska Läkartidningen* **15**, 1577–1600.

Philipson, T.: 2000, Economic epidemiology and infectious diseases, *Handbook of Health Economics* **1**, 1761–1799.

Pinheiro, P., Mathers, C. and Krämer, A.: 2010, *The global burden of infectious diseases*, Springer.

Price-Smith, A.: 2001, *The health of nations: infectious disease, environmental change, and their effects on national security and development*, Mit Press.

Puhani, P. and Sonderhof, K.: 2010, The effects of a sick pay reform on absence and on health-related outcomes, *Journal of Health Economics* **29**(2), 285–302.

Rauhut, D.: 2002, *Fattigvård, socialbidrag och synen på fattigdom i Sverige*, Lund Studies in Economic History 18, Lund.

Riksskatteverket: 2003, *Deklarationen 100 år och andra tillbakablickar: En jubileumsbok*, Norstedts juridik, Stockholm.

Robalino, D., Voetberg, A. and Picazo, O.: 2002, The macroeconomic impacts of AIDS in Kenya estimating optimal reduction targets for the HIV/AIDS incidence rate, *Journal of Policy Modeling* **24**(2), 195–218.

Rodrik, D. and Wacziarg, R.: 2005, Do democratic transitions produce bad economic outcomes?, *The American Economic Review* **95**(2), 50–55.

Roine, J. and Waldenström, D.: 2009, Top incomes in Sweden over the twentieth century, *Top Incomes: A Global Perspective* **2**.

Rose, A.: 2000, One money, one market: The effect of common currencies on trade, *Economic Policy* **15**(30), 7–46.

Rosling, H.: 2012, Total population for countries and territories.

Ross, R.: 1916, An application of the theory of probabilities to the study of a priori pathometry. part i, *Proceedings of the Royal Society of London. Series A* **92**(638), 204–230.

Ruhm, C.: 2000, Are recessions good for your health?, *The Quarterly Journal of Economics* **115**(2), 617–650.

Russell, P. et al.: 1955, *Man's Mastery of Malaria*, Geoffrey Cumberlege, Oxford University Press.

Ryberg-Welander, L.: 2000, *Arbetstidsregleringens utveckling: En studie av arbetstidsreglering i fyra länder*, Sociologiska institutionen, Stockholms Universitet.

Sachs, J. and Warner, A.: 1997, Sources of slow growth in African economies, *Journal of African Economies* **6**(3), 335.

Schön, L.: 2010, *Swedens's Road to Modernity: An Economic History*, SNS förlag.

Scoppa, V. and Vuri, D.: 2012, Absenteeism, unemployment and employment protection legislation: Evidence from italy.

Shapiro, C. and Stiglitz, J.: 1984, Equilibrium unemployment as a worker discipline device, *The American Economic Review* pp. 433–444.

Simonsen, L., Clarke, M., Schonberger, L., Arden, N., Cox, N. and Fukuda, K.: 1998, Pandemic versus epidemic influenza mortality: A pattern of changing age distribution, *Journal of Infectious Diseases* **178**(1), 53.

Smith, A.: 1776, *The Wealth of Nations*.

Statistics Sweden: 1907, *Bidrag till Sveriges Officiella Statistik. A) Befolkningsstatistik*, Stockholm: Kungl. Statistiska Centralbyrån.

Statistics Sweden: 1911, *Sveriges officiella statistik. Kommunernas Fattigvård och Finanser*, Stockholm: Kungl. Statistiska Centralbyrån.

Statistics Sweden: 1915, *Dödsorsaker år 1911*, Stockholm: Kungl. Statistiska Centralbyrån.

Statistics Sweden: 1917, *Folkräkningen den 31 December 1910*, Stockholm: Kungl. Statistiska Centralbyrån.

Statistics Sweden: 1921, *Skattetaxeringarna samt fördelningen av inkomst och förmögenhet*, Vol. 1917, Stockholm: Kungl. Statistiska Centralbyrån.

Statistics Sweden: 1926, *Folkräkningen den 31 December 1920*, Stockholm: Kungl. Statistiska Centralbyrån.

Statistics Sweden: 1936, *Folkräkningen den 31 December 1930*, Stockholm: Kungl. Statistiska Centralbyrån.

Statistik der Bundesagentur für Arbeit: 2013, Beschäftigungsstatistik, Sozialversicherungspflichtig Beschäftigte nach Altersgruppen zum Stichtag 30. Juni 2012, Nürnberg, Datenstand Januar 2013.

Statistisches Bundesamt: 2013a, Arbeitslosenquote aller zivilen Erwerbspersonen: Bundesländer, Jahre, Geschlecht, https://www-genesis.destatis.de/genesis/online, Date of retrieval: January 2013.

Statistisches Bundesamt: 2013b, Indikator 3.1 des Indikatorensatzes der GBE der Länder: Sterbefälle (Anzahl, je 100.000 Einwohner, je 100.000 Einwohner -altersstandardisiert an der Alten Europastandardbevölkerung-) nach Geschlecht, Region, ab 1998, www.gbe-bund.de , Date of retrieval: January 2013.

Stock, J. and Watson, M.: 2008, Heteroskedasticity-robust standard errors for fixed effects panel data regression, *Econometrica* **76**(1), 155–174.

Tabellini, G.: 2010, Culture and institutions: economic development in the regions of europe, *Journal of the European Economic Association* **8**(4), 677–716.

Taubenberger, J. and Morens, D.: 2006, 1918 influenza: The mother of all pandemics, *Rev Biomed* **17**, 69–79.

Tumpey, T., Basler, C., Aguilar, P., Zeng, H., Solórzano, A., Swayne, D., Cox, N., Katz, J., Taubenberger, J., Palese, P. et al.: 2005, Characterization of the reconstructed 1918 Spanish influenza pandemic virus, *Science* **310**(5745), 77.

Turkington, C. and Ashby, B.: 2007, *The encyclopedia of infectious diseases*, Infobase Publishing.

UN: 2008, Report on the global hiv/aids epidemic, Joint United Nations Programme on HIV/AIDS.

UN Population Division: 2004, *The Impact of AIDS*, Economic and social analysis of technology, United Nations.

UNAIDS: 2010, UNAIDS report on the global AIDS epidemic 2010.

United Nations: 2009, *Trends in International Migrant Stock: The 2008 Revision*, United Nations, Department of Economic and Social Affairs.

Watts, S.: 1999, *Epidemics and history: disease, power and imperialism*, Yale University Press.

Weber, J. and Courvalin, P.: 2005, An emptying quiver: antimicrobial drugs and resistance, *Emerg Infect Dis* **11**(6), 791–793.

Wooldridge, J.: 2009, *Introductory econometrics: A modern approach*, South-Western Publications.

World Health Organization: 2004, The world health report 2004: Changing history, annex table 2: Deaths by cause, sex and mortality stratum in who regions, estimates for 2002.

Young, A.: 2005, The gift of the dying: The tragedy of AIDS and the welfare of future African generations, *Quarterly Journal of Economics* **120**(2), 423–466.

Young, A.: 2007, In sorrow to bring forth children: fertility amidst the plague of HIV, *Journal of Economic Growth* **12**(4), 283–327.

Ziebarth, N. and Karlsson, M.: 2010, A natural experiment on sick pay cuts, sickness absence, and labor costs, *Journal of Public Economics* **94**(11), 1108–1122.

Zweifel, P., Breyer, F. and Kifmann, M.: 2009, *Health economics*, Springer.

Appendix

A Data Sources of Chapter 2

HIV data: Our primary data source is Oster (2007). Moreover we combine this with data from UN (2008). In order to minimize the HIV prevalence in our control group we only include countries in the control group where the maximum of these two measures in any year does not exceed one.

Total GDP, GDP per capita and population: Most of our data comes from the World Bank database (Mundial 2011). Here we combined GDP data from different available variables and transformed all data to purchasing power parity level in current dollar currency. For some countries (Alan Heston et al. 2006) had richer data on GDP per capita. Combining this dataset with a richer dataset for population from (Rosling 2012), we replace missing values in the original data both at the total level and in per capita terms.

Gross capital formation (total and per capita): Again the Development Indicators from World Bank are our main source Mundial (2011). Moreover we use gross capital formation data from United Nations Cheung (2008). Moreover we replaced missing values by dividing total capital formation by the number of inhabitants from (Rosling 2012).

Regime characteristics: Polity IV dataset (Marshall et al. 2009)

Political Violence: Major Episodes of Political Violence dataset (Marshall 2010)

Human Capital: This variable is measured in terms of year of schooling as provided by (Barro 2001). Moreover, we find high correlation between this measure and Lutz et al. (2007), who measured percent of population that completed primary, secondary and tertiary education respectively. Therefore

we replace missing values in the Barro dataset with data from Lutz transformed into years of schooling.

Labor force, percent of females in the labor force, arable land in hectares, percent of population living in rural areas and population growth(total and per capita): World Bank database development indicators (Mundial 2011). Again we replaced missing values by dividing total values by the number of inhabitants from (Rosling 2012)

Furthermore we use data from different papers from the empirical literature:

Climate and soil quality and data on physical geography (e.g. average meters above sea level) and population from Gallup et al. (2001)

Ethnic and cultural fractionalization from Fearon (2003)

Religion, influences from colonialization, life expectancy, fertility and other determinants of economic growth from Ciccone and Jarocinski (2010)

Quality of the government along several dimensions such as bureaucratic delays, the estimated size of the black market, the level of corruption and other variables from La Porta et al. (1999) and Englebert (2002)

Globalization index provided by Dreher (2006)

Ratio of GDP to GNP from Alan Heston et al. (2006)

B Data Sources for Chapter 3

HIV data: Our primary data source is Oster (2007). Moreover we combine this with data from UN (2008). In order to minimize the HIV prevalence in our control group we only include countries in the control group where the maximum of these two measures in any year does not exceed one.

Birth rate, death rate, fertility rate and life expectancy: World Bank database development indicators (Mundial 2011). Again we replaced missing values by dividing total values by the number of inhabitants from (Rosling 2012)

Human Capital: This variable is measured in terms of year of schooling as provided by (Barro 2001). Moreover, we find high correlation between this measure and Lutz et al. (2007), who measured percent of population that completed primary, secondary and tertiary education respectively. Therefore we replace missing values in the Barro dataset with data from Lutz transformed into years of schooling.

Total GDP, GDP per capita and population: Most of our data comes from the World Bank database (Mundial 2011). Here we combined GDP data from different available variables and transformed all data to purchasing power parity level in current dollar currency. For some countries (Alan Heston et al. 2006) had richer data on GDP per capita. Combining this dataset with a richer dataset for population from (Rosling 2012), we replace missing values in the original data both at the total level and in per capita terms.

C Alternative Specifications (Chapter 4)

Table C-1: Regression Results. Alternative Specifications.

	Outcome Variable in Levels					Results for Super-Regions				
Panel A: Capital income	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	-35.08	-10.84		-18.17	-8.139	-1.246**	-1.023		-1.086*	-1.001
	34.08	28.71		38.49	36.12	0.345	0.667		0.538	0.703
$w_{it} \times \mathbf{1}(t > 1920)$		-28.35**			-26.39**		-0.253			-0.198
		11.21			11.55		0.375			0.404
Placebo ($w_{i,t+3}$)			1.016					-0.112		
			5.043					0.196		
Constant	27.89***	27.89***	27.89***	27.81***	27.81***	3.136***	3.136***	3.136***	3.179***	3.179***
	3.958	3.962	1.492	3.742	3.761	0.0503	0.0506	0.0334	0.0474	0.0486
Further Controls	No	No	No	No	No	No	No	No	No	No
Regional Time Trends	No	No	No	No	No	No	No	No	No	No
Observations	475	475	150	125	125	114	114	36	30	30
R^2	0.603	0.604	0.318	0.595	0.597	0.983	0.983	0.885	0.979	0.979
Panel B: Earnings	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	-49.21	-52.92		-41.57	-44.05	-0.327	-0.303		-0.376	-0.371
	121.4	55.36		96.96	60.96	0.185	0.321		0.207	0.282
$w_{it} \times \mathbf{1}(t > 1920)$		4.341			6.539		-0.0276			-0.0108
		114.4			124.5		0.180			0.185
Placebo ($w_{i,t+3}$)			-11.26					-0.0131		
			37.15					0.177		
Constant	301.5***	301.5***	301.5***	318.2***	318.2***	5.634***	5.634***	5.634***	5.714***	5.714***
	10.59	10.60	11.90	6.372	6.406	0.0399	0.0401	0.0311	0.0208	0.0213
Further Controls	No	No	No	No	No	No	No	No	No	No
Regional Time Trends	No	No	No	No	No	No	No	No	No	No
Observations	475	475	150	125	125	114	114	36	30	30
R^2	0.836	0.836	0.357	0.827	0.827	0.989	0.989	0.740	0.986	0.986
Panel C: Poverty	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
w_{it}	3.719***	0.143		1.782*	0.217	0.967***	0.131		0.587**	0.193
	1.203	0.526		0.884	0.500	0.172	0.178		0.166	0.158
$w_{it} \times \mathbf{1}(t > 1920)$		4.183***			4.119***		0.948***			0.914***
		0.937			1.040		0.0432			0.0705
Placebo ($w_{i,t+3}$)			0.230					0.0508		
			0.404					0.115		
Constant	4.393***	4.393***	4.393***	4.453***	4.453***	0.775***	0.775***	0.775***	0.794***	0.794***
	0.151	0.151	0.0505	0.127	0.128	0.0266	0.0268	0.0179	0.0205	0.0207
Further Controls	No	No	No	No	No	No	No	No	No	No
Regional Time Trends	No	No	No	No	No	No	No	No	No	No
Observations	475	475	150	125	125	114	114	36	30	30
R^2	0.487	0.512	0.150	0.468	0.545	0.803	0.832	0.262	0.783	0.867

The table shows three panels with results from fixed effects regressions. In the first specification we regress the dependent variable (which varies with each panel) on our treatment variable w_{it}. The second column additionally interacts the treatment variable with a dummy which equals one for data points after the Spanish Flu. The third column presents the placebo regression which is estimated using years before 1918 and the third lead of the treatment variable ($w_{i,t+3}$). The final specifications 4 and 5 collapse the data in order to control for autocorrelation. Columns 6–10 present exactly the same specifications but for the logarithmic outcome variables. In these additional specifications we have collapsed the 25 counties into six 'super-regions' with approximately one million inhabitants each. All regressions include year dummies which are not displayed and are weighted by the population in 1917. The second row presents robust standard errors. As discussed by Wooldridge (2009), Stock and Watson (2008) and Arellano (2003), in a fixed effects model, robust standard errors are equivalent to regionally clustered standard errors. The asterisks represent significance at the following p values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Proofs and Tables (Chapter 5)

Proof of Proposition 1

We calculate the conditions for stability of the fixpoints to identify whether in a neighborhood around the fixpoints attraction or repellence occurs. For this we first calculate the slope of the sequence at the fixpoints which equals:

$$f'(p^*_{\pm}) = 1 \pm \sqrt{4iqp_{ex}(1-p_{ex}) + (1-iq(1-p_{ex}))^2}. \tag{6.1}$$

For stable fixpoints we must have $|f'(p^*_{\pm})| < 1$. Therefore the first fixpoint (p^*_+) will be always unstable since the derivative is greater or equal to one.

Finally, considering the solution for $p_{ex} > 0$ and restricting the sample to stable fixpoints with $f'(p^*_-) > -1$, non-negative sickness probabilities $(p_t, i, q \geq 0)$ and a higher sick leave during recessions $p_t(1-q) > p_{ex}$ as found in the empirical literature, we obtain the following inequalities:

$$\left(p_{ex} = 0 \text{ and } \frac{1}{q} < i < \frac{3}{q}\right) \text{ or} \tag{6.2}$$
$$\left(0 < p_{ex} < \frac{1-q^2}{1+2q} \text{ and } \frac{1-q}{(1-p_{ex})^2 - q(1-p_{ex})} < i < \frac{2\sqrt{1-p_{ex}+p_{ex}^2}+1-2p_{ex}}{q(1-p_{ex})}\right)$$

■

Tables

Table D-1: Sickness categories based on 10th revision of International Classification of Diseases (ICD-10)

Category	Disease	Total Spells 2010	Total Days 2010	Group
A00–B99:	Certain infectious and parasitic diseases	553,546	3,155,185	2
C00–D48:	Neoplasms (Cancer)	80,466	3,061,937	1
E00–E90:	Endocrine, nutritional and metabolic diseases	29,197	552,346	1
F00–F99:	Mental and behavioral disorders	268,300	10,182,958	1
G00–H95:	Diseases of the nervous system, the eye and the ear	261,876	3,256,858	1
I00–I99:	Diseases of the circulatory system	166,238	3,581,913	3
J00–J99:	Diseases of the respiratory system	1,657,966	11,025,228	1
K00–K93:	Diseases of the digestive system	692,823	4,489,656	1
L00–L99:	Diseases of the skin and subcutaneous tissue	81,112	1,040,655	1
M00–M99:	Diseases of the musculoskeletal system and connective tissue	1,037,963	21,315,233	1
N00–N99:	Diseases of the genitourinary system	135,265	1,420,881	1
O00–O99	Pregnancy, childbirth and the puerperium	47,025	628,066	1
R00–R99:	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	299,819	3,077,885	1
S00–T98:	Injury, poisoning and certain other consequences of external causes	519,493	10,539,770	1
Z00–Z99	Factors influencing health status and contact with health services	81,792	1,689,116	1
	Total	5,912,881	79,017,687	
	Resulting from Workplace Accidents	170,224	3,602,501	

The table shows the categories for the diseases of our data source, with additional workplace accidents (already included in the previous cases). Moreover we report the total number of started sickness spells in 2010 for Germany. In the final column we show how we grouped the categories into three groups for the regression. These groups are non-infectious diseases (1), infectious diseases (2) and diseases of the respiratory system (3).

Hiermit erkläre ich, dass ich die Arbeit - abgesehen von den in ihr ausdrücklich genannten Hilfen - selbstständig verfasst habe.

Stefan Pichler

Werdegang

Nach Erlangung der Matura (Allgemeine Hochschulreife in Italien) an der Handelsoberschule in Bozen/Italien im Jahre 2004 hat Stefan Pichler von 2004–2007 an der Freien Universität Bozen den Bachelerstudiengang Economics and Management absolviert. Dieses Studium hat er mit voller Punktezahl (110/110) abgeschlossen.

Anschließend besuchte er das Masterprogramm in Quantitative Economics an der Goethe Universität Frankfurt, welches er 2010 abschloss. Von 2009-2011 arbeitete er als Wissenschaftlicher Mitarbeiter am Fachgebiet für Angewandte Mikroökonomik und Insitutionenökonomik an der Technischen Universität Darmstadt. Seit 2011 ist er am am Fachgebiet Finanzwissenschaft und Wirtschaftspolitik der Technischen Universität Darmstadt als Wissenschaftlicher Mitarbeiter tätig. Außerdem ist er seit 2007 Student im PhD Programm der Graduate School of Economics, Finance and Management, einer Graduiertenschule, welche eine Kooperation der Universitäten von Darmstadt, Frankfurt und Mainz darstellt.