The Swedish payroll tax reduction for young workers

- A study of effects found using publicly available aggregated (macro) data

Balder Bergström

Student
2019
Master I, 15 ECTS
Master’s Programme in Economics
Abstract

In 2007, the Swedish payroll tax was reduced for youths in an attempt to suppress the perceived high unemployment among Swedish youths. The reform was rolled back later in 2016. For this period there is a rich supply of publicly available aggregated (macro) data. This thesis aims to examine: first, if the aggregated data is suitable for policy evaluation of the reform, and second, the effects of the reform introduction and repeal. This has been done by using both a conventional fixed effects model and a more unorthodox synthetic control method. Neither of the two methods could show any unbiased and consistent significant result of the treatment effects of the reform. Instead, the results of this thesis suggest that the publicly available aggregated data doesn’t contain enough information to evaluate such reforms.

Keywords: Payroll tax reduction, labor economics, labour economics, aggregated data, synthetic control method, arbetsgivaravgifter, arbetsmarknadsekonomi.
# Table of contents

1. Introduction......................................................................................................................... 1

2. Background and institutional settings................................................................................ 3
   2.1 The Swedish Payroll tax .................................................................................................. 5
   2.2 Reform and repeal ......................................................................................................... 7
   2.3 The wage setting in Sweden. .......................................................................................... 9

3. Payroll tax reduction in economic theory ........................................................................... 11
   3.1 The general case (a reduction for all workers) .............................................................. 11
   3.2 The targeted case (a reduction for a specific group)....................................................... 12

4. Litterateur review .............................................................................................................. 15

5. Empirical approach ........................................................................................................... 20
   5.1 Common trend assumption and public data ................................................................. 21
   5.2 Data ............................................................................................................................... 23
   5.3 The least square approach ............................................................................................ 25
   5.4 The synthetic control group approach .......................................................................... 30

6. Estimation and Results ....................................................................................................... 34
   6.1 Results from the fixed effects model ............................................................................ 35
   6.2 Results from the synthetic control method ................................................................... 36

7. Discussion .......................................................................................................................... 39

References .................................................................................................................................. 41

APPENDIX .................................................................................................................................. 44
1. Introduction

During the first years of the 21st century, the unemployment rate among Swedish youths increased rapidly. This led to a broad and intensive political debate. During the summer of 2007 the newly elected center-right government introduced a vast payroll tax reduction, with retained benefits, for all young workers to manage this labor market problem. In 2009, the cut was further extended so that the payroll tax rate on young workers was half of the normal payroll tax rate. The reform was presented as permanent, but when the earlier center-left opposition was elected into office they announced the repeal of the reform. In 2016 the payroll tax rates were the same for all age groups and the reform was fully repealed.

Since the core in economics is to optimize the allocation of resources, it is of interest to evaluate the effectivity of this long-lasting policy reform. The idea of handling the problems of specific groups on the labor market by reducing the hiring cost of this group is not a specific Swedish idea, but rather a quite common proposal. Hence an evaluation of this reform may serve more than only Swedish policymakers. Although the short-term introduction effect of this reform has already been studied by several researchers, to the best of my knowledge, no one has studied the effect of the withdrawal of the payroll-tax reduction. Since there may be other effects of labor cost reduction than labor cost increases, there is a gap in this research field to be filled.

Additionally, there is a substantial amount of aggregated data that can be obtained from public databases. Since many political interventions of interest for evaluation, as for example the one evaluated in this thesis, take place on an aggregated level, such data has large potential for evaluation purposes. Most research today is done with the use of microdata were individuals are observed. Such data is costly to collect. If these assessments could be based on the already publicly presented data, it would make policy evaluation possible for more than only the researchers and institutions with large budgets.

With this background it is also of interest for the field of economics to test whether publicly published data can generate enough information to be the foundation of a study. Since this intervention was on an aggregated level, and numerous independent researchers have presented estimates of the treatment effect for the introduction of the reform that be used for comparison, it is possible to test whether or not the publicly presented aggregated data is good enough for policy evaluation. Thus, the research question of this thesis is:
Is it possible to obtain an unbiased and consistent estimate of the treatment effect of the introduction and repeal of the reduced payroll tax for young workers in Sweden only using publicly available aggregated (macro) data?

To answer this research question two different estimation methods have been used, both a conventional fixed effects model and the more novel synthetic control method. These two estimation methods have both been applied on various identical datasets, originating from different public data sources, describing the employment and unemployment rates among young adults\(^1\). This will be further presented in the following sections of this thesis which are organized as follows: Section 2 discusses the background to the reforms and the features on the Swedish labor market; the expected outcome of a payroll tax subsidies in general and targeted on a specific group, based on economic theory, will be briefly considered in section 3; section 4 provides literature with a focus on previous empirical evaluations on this specific reform; The data and empirical approach of this study are introduced in section 5; and in section 6 the results from this thesis will be presented, followed by a brief discussion about the results in section 7.

---

\(^1\) The employment/unemployment among European youths both on countries and regional level and the employment/unemployment among different age cohorts in Sweden. However, on the later dataset the synthetic control method wasn’t feasible.
2. Background and institutional settings

Labor is the path to wealth for both the individual as well as for the State. Taxes on labor stands for approximately 60 percent of Sweden’s total tax revenue (Government bill 2018/19:1) and may be regarded as the foundation for not only all tax revenue but also all wealth. Thus, a high employment rate is the foundation for high wealth both for the welfare state, as well as for common citizens. It is, therefore, of great interest for the policymakers to create such a labor market setting that the employment rate is held high and none is involuntarily unemployed. Sweden has had, since the millennium, an employment rate among working-age population about 10 percentage points above the OECD average (OECD, 2019), and, during the last decade, on average, the third highest within the EU (IFAU, 2017:15).

A high employment rate among the population can be explained by either a high labor force participation or a low unemployment rate within the labor force, or a combination of both. The main explanation of differences in the employment rate among the European countries is the differences in labor force participation, i.e. that people can and have the will to work when they have the opportunity. This is also the source of Sweden’s high employment rate. No other EU country has a higher labor force participation rate. The explanation for this is mainly high participation among women and elderly people (IFAU, 2017:15).

A person who participates in the labor force but who isn’t employed, i.e. he or she is searching for a job, is defined as unemployed. Although the unemployment rate in Sweden has been fluctuating around the EU average, Sweden is compared both to historic data and other northern European countries at rather unflattering rates. When sorting unemployment data by age, one can clearly see that the young labor force participants have the highest unemployment rate among the age groups. Even though Sweden also for this age cohort\(^2\) has had an unemployment rate fluctuating around the European average (see figure 1 and 3), Sweden is clearly above some other northern neighbors (see figure 4).

---

\(^2\)Youths is in European comparative statistics defined as either the age cohorts 15-24 or 20-24. In this thesis I try to use 20-24 as much as possible since it is closest to the age groups targeted by the payroll tax cut. Depending on which definition used, Sweden’s position in international comparison differs.
Figure 1 Unemployment rate

Y-axis describes the unemployment rate among the different groups. *EU is average of all European countries with missing value for maximally one year. Source: Eurostat.

Figure 2 Students among unemployed

Unemployment rate among Swedes in age 15-24 based on whether or not they are full-time students. Data is missing pre second quarter 2005. Source: SCB.

Figure 3 Unemployment a European comparison

Time average (period: 2001-2017) unemployment rate among the European countries, (see which above), divided into quantiles. Two age cohorts showed. Sweden presented separately. Source: Eurostat.

Figure 4 Comparison Northern countries

Time average (period 2001-2017) NEET-rate, unemployment-rate, and employment-rate among six northern European countries and Sweden. All presented separately. NEET-rate is for age cohort 15-24 whereas both other for age cohort 20-24. Observe that the employment-rate in percent is to the right. Source: Eurostat.

It is clear that the unemployment rate among Swedish young adults was, during the first years of this millennium, steeply increasing. The same pattern can’t be seen in youth employment rates (see figure 7 later on in paper) or the youth NEET-rate\(^3\), neither rising more than marginally during the same time period. Since the Swedish NEET-rate is fairly low, one must assume that the vast majority of Swedish youths are occupied by either studies or employment. One possible explanation for the ambiguous tendency at the beginning of the millennium, with

\(^3\) NEET is the acronym for “neither in employment nor in education and training”.
rising youth unemployment but constant youth unemployment rate and NEET-rate, can be an increase in students searching for jobs.

Every unemployed respondent who is searching for a job, independent of whether or not he or she is a full-time student, is registered as unemployed. The increase could thus simply be an increase of voluntary unemployed youths that in fact are occupied and are searching for a side job or an attractive alternative to their full-time studies. This interpretation is discussed by Eriksson, Hensvik, and Nordström Skans (2017). It is, however, hard to find public data that confirms or discards this hypothesis, in figure 2 one can, however, see that about half of the unemployed in age cohort 15-24 are in fact, full-time students.

That said, the purpose of the reform, which is evaluated in this thesis, was to reduce the youth unemployment that was at the time perceived in the political debate to be both rising far too rapidly and being excessively high. Youth unemployment has, in fact, been the dominating topic in the post-millennium political discussion about the labor market in Sweden until recently when refugee integration has seized the baton.

### 2.1 The Swedish Payroll tax

In Sweden there is a compulsory linear payroll tax rate of 31.42 percent paid by the employer on top of the worker’s salary, i.e. the gross cost of labor is 131.42 percent of the gross wage. The payroll tax includes fees for six social security benefits conditional on labor force participation including parental insurance, disability insurance, widow insurance, health insurance, labor market fee, pensions, and the general wage fee. The general wage fee isn’t connected to any benefits and is, therefore, more similar to a tax than a fee, even though there is a diffuse borderline separating a tax from a fee for social benefits (Government bill 2014/15:1). The payroll tax rate has been quite consistent over time but was in 2009 reduced from 32.42 percent to the current rate of 31.42 percent. During the last couple of years, there has been a shift in the distribution towards an increased general wage fee at the expense of especially health insurance (Skatteverket, 2008:2019). This shift is clearly described in figure 5 below.
What should be emphasized is that none of the fees, except for pensions, are directly connected to the benefit itself. Neither in possibility to take part of the benefit nor in financing the cost of the social benefit system, i.e. even though the value of the collected fee for each separate social benefit is quite balanced to the actual cost of each social benefit, the revenue from the separate fee within the payroll tax isn’t earmarked to finance the specific benefit (Government bill 2015/16:1; 2018/19:1). Instead, the collected money from the fees goes into the government budget and the social benefit is financed by the government budget. This gives the government the possibility of using other income to finance the costs when they exceed the income of the fees and vice versa when so is the case, the latter being more common (Government bill 2015/16:1; 2018/19:1).

Pensions are, contrary to the other fees, directly connected both in possibility to take part in the benefit and in financing the benefit’s system. It’s a self-financed system on the side of the government budget. Pension is based on the pension justified income (gross income from labor and in some cases subsidies as parental, study, etc.). More specifically, in addition to the net income, every entitled person gets 18.5 percent of the pension justified income paid into their
public pension account. The funding of the pension account can be deduced to two inflow channels. 10.21 percentage points come from the payroll tax and the rest is deducted from the income tax (Government bill 2014/15:1).

2.2 Reform and repeal

In 2007 the newly elected center-right government adopted, among many other reforms, a payroll tax cut targeted to young workers, as the previously described youth unemployment was a big topic in the political debate at the time. The motive for the reform was to ease the labor market entry for young adults. By reducing the labor cost for young adults, the demand for them was intended to increase. The explicit reform was a reduction by 50 percent of all components of the payroll tax, excluding pension, for workers which during the calendar year turned at least 19 but not 26 years. The reason for the lower age threshold was that the government didn’t want to create any incentives for youths to exit upper secondary school. This led to a reduction of the payroll tax of 11.1 percentage points and correspondingly an 8.4 percent reduction of labor cost for firms. The reform was implemented on the first of July, 2007 (Government bill 2006/07:84).

On the first of January, 2009, an extension, in both age range for those who were treated and extent of the treatment, gained legal force. All components of the payroll tax, except pension, were further reduced to 25 percent of the general value, implying a reduction of the payroll tax of 15.9 percentage points and of labor cost for firms with 12 percent correspondingly, when comparing to the original payroll tax rate. The higher age threshold was elevated. The lower age threshold was removed, with the motive to simplify administration and, more importantly, create a greater demand for summer jobs during the school vacation. These pros were considered greater than the cons earlier described.

The motivation for increasing both the upper age threshold and the subsidies was that there was a need for a further endeavour to simplify the entrance on the labor market for youths (Government bill 2008/09:7). One crucial point is that neither of these reforms was affecting the workers benefits, even though the fees were only partly paid. Further, it is important to note that the administration for firms to register for subsidies was minimal. This, in combination

---

4 Since the Swedish pension system mainly is a Pay-as-you-go system the vast majority (16 percentage points) of the payment into the system is converted to pension rights, a right to realize the savings in the future and only 2.5 percentage points is actual placed in a pension foundation account. Remember that the full payment into the system is 18.5 percentage points of the pension justified income.
with widespread awareness of reform made, the take-up among firms almost perfect, according to Saez, Schoefer, and Seim (2017). They argue however that the political disagreement among parties in the parliament about the reform may have reduced the effect of the reform.

Figure 6 The timeline of the payroll tax reduction

Y-axis describes the payroll tax in percentage of the gross wage. The four vertical lines describe in order from left: 1. The introduction of the reform. 2. The extension of the reform. 3. The first step of the repeal. 4. The full repeal of the reform.

In 2014 a new center-left government was elected. These parties had, during years in opposition, been critical to the reform and in the budget proposal for 2015, the repeal of the reform was advertised. According to the new government, the reform was too ineffective, since also already occupied youths were subsidized the deadweight loss was substantial (Government bill 2014/15:1). The repeal was made in two steps in order to ease the monetary loss for firms associated with withdrawal. On the first of May, 2015, the first step was implemented. Workers who, during the calendar year, turned at most 23 years old got a full subsidy for all fees except the pensions. Whereas for workers who will turn 24, but not more than 25 was the subsidies remained the same. For workers above this threshold the subsidies were repealed, i.e. the upper threshold was reduced by two years (Government bill 2014/15:50). As 2015 became 2016 the full repeal was in place and every employer had to pay full payroll tax for all workers regardless of age.
2.3 Wage setting in Sweden.

According to economic theory, which will be further presented in the next chapter, the effect of a payroll tax for a targeted group depends a lot on the wage setting system. This urges for a brief overhaul of how wages are set in Sweden. In comparison to most other OECD countries, Sweden has no national minimum wage prescribed by law. Instead, the collective bargain agreements (CBA) sets the minimum wage. The minimum wage is not general but does instead differ quite substantially both between and within CBAs. Within a CBA the minimum wages are based on things as age, education, experience and tenure. According to Medlingsinstitutet (2015), approximately 90 percent of the workers were covered by a CBA for the years 2007-2015.

Every blue-collar sector has generally one trade union and one employer’s organization that regularly bargains a CBA, usually every second year. The fourteen trade unions for blue-collar workers are affiliated to the umbrella organization “Landsorganisationen i Sverige” (LO), in English “The Swedish trade union confederation”⁶. In 2007 around 74 percent of all blue-collar workers were members in a trade union, and an absolute majority in one of the LO-affiliated unions (Kjellber, 2011). The employer’s organizations, on the other hand, are united in “Svenskt Näringsliv”, in English “The Confederation of Swedish Enterprises”. For blue-collar workers, the collective bargain agreements are set on industry level between the trade union and employer’s organization in every sector. However, Fredrikson, and Topel (2010) did conclude that most of the wage setting was determined on a local level. The central CBAs provided the frame for the agreeable increase in total labor cost on firm-level and the allocation of the increases was determined in local negotiations. The wage increase, or total labor cost increase, in all central agreements, are generally determined by the “Märke” (in English freely translated to “the Trace”). The “Märke” is the percent wage increase settled in the bargaining agreement made in the highly international competitive export sector, i.e. the CBA between trade unions and employers’ organizations within these sectors.

⁵ There are also blue-collar trade unions for specific occupations such as the Building workers union, the Electricians Union and the Painters Union, which could be said to be in the construction sector.
⁶ There are some small unions for blue-collar workers who are not affiliates to LO. The most commonly known is The Swedish Dock Workers Union. But these unions cover only a marginal part of the blue-collar workers.
Since most of the employees in the age cohorts affected by reform were blue-collar workers, their wages were in general decided through the minimum wage in their sector’s CBA, where factors such as age, education, experience, and tenure determine the wage. Although most of the payroll-tax-reduction-eligible young workers were blue-collar workers, one must assume that some were white-collar workers. The wage setting for these employees is far more decentralized, and normally no minimum wages are agreed in these sector’s CBAs. However, norms about suitable salary based on age and tenure, similar to the ones in blue-collar industries, still play a substantial role in the wage setting for white-collar workers.
3 Payroll tax reduction in economic theory

The relation between demand and supply decides the equilibrium of a market. The labor market is no exception. Employees demand labor which workers supply. When a reform such as the payroll tax subsidies is introduced the rules of the market are altered, creating a new equilibrium. But where this new equilibrium is located, i.e. how effective the reform is in terms of decreasing unemployment, depends in standard economic theory on to what extent the windfall money is levied on the employers or on the employees (Skeding, 2014).

It is of great importance to emphasize the difference between a cost reduction due to abatement of minimum wage and a fiscal policy funded rebate of the payroll tax, with retained benefits for workers. The first reduces the labor supply whereas the second doesn’t. A reduction of minimum wage contracts the difference between the reservation salary and the net wage, i.e. the utility from paid work is reduced in general. This will reduce the labor supply, since the labor force participants at the margin will gain more utility from leisure than paid work. A reduction of the payroll tax rate, with retained benefits, does, in conformity with a reduction of minimum wage, reduce labor cost for firms. It does not however reduce the net wage for workers nor the social security benefits. Since the relative utility gained from labor is unaltered, the labor supply isn’t affected by such reform.

3.1 The general case (a reduction for all workers)

Even though the supply of labor is not affected by a payroll tax subsidy, the demand for labor is, according to economic theory. As the cost of hiring labor decreases, the demand for labor increases, i.e. the demand curve shifts to the right. According to the standard textbook, this will generate two effects. Either the wage ($W$) or the employment ($L$) will increase, presumably both effects will occur. To which extent the windfall money is used to increase either wage or employment depends on the elasticity of labor supply ($\varepsilon$) and labor demand ($\eta$). The greater elasticity of labor supply, the more of the windfall money ($S$) will be levied on the employers, the greater employment effects and the more modest wage increases and vice versa. Greater elasticity of labor demand implies both larger wage and employment effects of the payroll tax subsidy (Lawrence,1996). This is visualized in equation (1) and (2) on the next side.
\[
\frac{\partial \ln L}{\partial S} = \frac{\eta \varepsilon}{\eta + \varepsilon}
\]

\[
\frac{\partial \ln W}{\partial S} = \frac{\eta}{\eta + \varepsilon}
\]

In the short run, wages are rigid. This is a consequence of the CBAs, described in the earlier chapter, fixing the wage levels for a few years. Hence the labor supply is nearly perfectly elastic in the short run. The employment effects of a payroll tax reduction should, therefore, be quite extensive in the short run. Since the windfall money generated by reform is levied to the employers they should, according to theory, hire more labor (Skedinger, 2014). What happens, in the long run, however, is more of an enigma.\(^7\) As the trade unions re-bargain their CBAs one can assume that the windfall money generated through reform will at least partially be levied to the workers in wage increases. At which extent is however unclear and is probably affected by both labor supply elasticity, bargain power and the trade unions’ view on the trade-off between insiders and outsiders. Since the wage setting in an international context is very centralized in Sweden, it is suggested by Calmfors and Driffill (1988) that the trade unions will take outsiders into account. If so, at least some of the windfall money will be levied to the firms, giving them the possibility to increase employment.

3.2 The targeted case (a reduction for a specific group)

The payroll tax reduction introduced in 2007, and extended in 2009, was, however, not a general reduction treating every worker on the labor market. The intended effect of the reform was to reduce youth unemployment and thus the subsidy of the payroll tax was only targeted towards young workers. Therefore, the relative cost of labor is of great interest. When assuming the productivity of workers increases with age, at least to some certain limit (Saez, Schoefer, and Seim, 2017), then the firm wants to hire a younger worker rather than a more senior worker, only if the labor cost for the younger is lower than for the more senior. The firm is indifferent between the two when the productivity per krona is equal. If we assume that the labor market was a decentralized perfectly working competitive market with rational agents, the wages would have been set at such levels initially.

\(^7\) This can however be said to not matter that much since according to Keynes (1923) “In the long run we are all dead”. 12
If the labor market was in equilibrium before the intervention. Then, as the reform was implemented, the eligible young workers would at the margin offer more productivity per krona compared with older workers, leading to that firms ought to substitute older workers to the more cost-effective younger. In the short run, it would increase employment among youths. In the longer run, the effects depend on the settings on the labor market. In the decentralized economy, the increased demand would in addition to increasing employment among youths, push up the net wage for the coveted young workers until the productivity per krona once again was equal between the age groups (Saez, Schoefer, and Seim, 2017).

The settings on the Swedish labor market is however, as earlier described, quite far from decentralized. Due to that the Swedish wage setting system is so rigid and norm-based, it can be suggested that the wages for the eligible young workers remain rather constant in comparison to the more senior workers. When the members of the trade unions have such substantial power in the allocation of the agreed labor cost increase, it is not plausible to assume that they choose to give this only to the for-reform-entitled young workers. That would be totally contractionary to the wage setting norms, where for example age and experience have had a positive effect on the wage. If instead assuming that the workers will share the rent of the reform, the new relative price of labor will be fairly persistent. Hence firms will hire more youths, until reaching equilibrium. Remember that equilibrium is where the productivity per krona is equal between the age groups. This builds on the assumption that firms start with hiring the most productive labor.

It should be stressed that whom the windfall money is levied to still matters for the size of the effect. In the case where all windfall money generated from reform is levied to the workers, in terms of higher wages for all workers, the increase in employment for youths will be at the expense of older workers and only due to the change in the relative price of labor. If the windfall money, at least to some extent, ends up in the hands of the employer’s, the firms may still substitute staff, but also increase its staff. Hence if some of the windfall money is levied to the employer, the effect on employment will not only be greater, but also not as much in disfavor of the unentitled workers, though this argumentation is founded on the assumption that firms will use the windfall money to hire labor. Such an assumption might be somewhat strong when both profits and investments might be attractive for firms. An additional component effecting the willingness to substitute more senior employees towards younger who are eligible for subsidies is the risk aversion. There is always a risk associated with changing staff, and an extra high risk-premium of hiring a young, untrained and unexperienced worker (Egebark and
Kaunitz, 2017). If the expected cost of this risk is too high, no substitution will be made. To summarize the theory, one can conclude that it is likely to see a short-run effect on employment. In the long-run, it is far more uncertain what will be the result of this reform.
4 Literature review

This specific payroll tax reduction has previously been studied and this thesis isn’t unique in its ambition to evaluate the effects of the reduction of payroll tax for young workers. If this chapter were to discuss studies of similar cases where the labor cost for a specific group is reduced but the gross income plus social benefits are retained for the worker, it would be ridiculously long. Hence only the most influential and relevant studies will be discussed. Two things uniting the studies here will be reviewed: firstly, that the data is microdata, meaning that the observations are individuals and not aggregated units; and second, they all rely on a conventional difference in difference approach (DiD), which suits microdata but might not suit aggregated data as well.

During the nineties, there were large fluctuations in the employer-paid tax contributions, i.e. payroll tax, for employees in the lowest wage segment in France. During the same period, the minimum wages were continuously increasing. This fact was exploited in an article by Kramarz and Philippon (2001) where they studied the effect of both labor cost increases and decreases in labor demand. The authors could show a labor demand elasticity of -1.5 on a labor cost increase. However, they could not find any significant connection between decreases in labor demand and cuts in the hiring cost of labor.

Both Sweden and Finland have, during the past decade, tried subsidies in payroll tax for rural regions. In Sweden, the reduction was nearly twice as big as in Finland, cutting the gross labor hiring cost by 7.3 percent and 3.3 percent respectively. Both reforms included a ceiling level of subsidies to a single firm. The Swedish payroll tax subsidies for the rural region were studied by Bennmarker, Mellander, and Öckert (2009). The results showed no economically significant effects of the policy reform, neither on wage nor employment rate. The Finnish reform was studied in an article by Korkeamäki and Uusitalo (2009), using other rural regions not as rural as the treated areas as control group they could conclude that no significant results or clear pattern were to find. The authors also supplied an estimation on hourly wages based on a subgroup from which they had more detailed income information. The result of this evaluation suggested a significant increase in wages for employees in treated firms. The wage increase took approximately half of the cost reduction, i.e. the windfall money. However, they could not find any evidence showing that the rest of the cost reduction was used by the firms to hire more workers and thereby increase employment levels.
The most influential study, and referred to in political debates\(^8\), on the effects of the Swedish payroll tax reduction, i.e. the reform which this thesis is evaluating, is without exception the one made by Egebark and Kaunitz (2013). In 2017 they extended the same paper to include more years and a more rigorous discussion. They investigated the employment effects on the affected group by using the age group as close to, but above, the higher threshold as control group and adding several individual-specific covariates that affect the probability of being employed. A modest but highly significant treatment effect of an employment increase of approximately around 2.7 percent in 2007 and 1.4 percent in 2008 was found. Based on results they estimate that the labor demand elasticity for young workers to be -0.31.

In addition to expanding the time-series Egebark and Kaunitz (2017) investigated some other things of interest. For instance, how the estimates vary depending on the definition of employment. In the original paper a person was defined as employed if he or she received a yearly income from labor about 25 percent of what a full-time job would generate. By setting the lower limit of employment to be a half-time or full-time employment the estimates were reduced and insignificant. If on the other hand the lower limit was relaxed to an even lower level, the estimates weren’t increased substantially. The authors also argued that they had findings showing that there was no persistence of the treatment effect (lagged effect).

Skedinger (2014) investigated the impact on employment and wage for the treated group in the retail industry as well as the profit effect on the firms in the sector depending on the degree of young employees within the staff. The methods used for evaluating the effect on employment and wage are quite similar to the one used by Egebark and Kaunitz, however, there are features separating the methodology between the studies, also aside from data restriction generated differences. Skedinger doesn’t use employment as the dependent variable. Instead, he uses working hours, new hiring and separation respectively. By substitution, Skedinger could count the net effect of occupation.

---

\(^8\) It is even referred to in the reform bill for the withdrawal of the reform.
The results show an insignificant increase in employment in the short run. In the longer run, however, the results are a significant and equal increase of both entry and exit, implying no net effect of employment, just a higher velocity of occupations. Estimates of the treatment effect of weekly working hours and hourly wages don’t show any significant results for blue-collar workers, except that the initial salary for new employees, in the long run, shows a significant, but minor, increase. Although Skedinger uses a rigorous micro dataset for these tests the parallel trend assumption is violated in the placebo treatment test, casting some doubts on the regression estimates. The estimation of profit effect for firms shows a significant positive effect indicating increased profit as the amount of for-subsidy-entitled employees increase. However, due to specification of the variables, it is however not possible to compare it with the amount of windfall money received by firms due to reform in a meaningful way.

Saez, Schoefer, and Seim (2017) pioneered the view of visualizing the firms, and not the individuals in the targeted age cohorts, as the treated group. The authors conclude that previous papers that examine the effect of the payroll tax reduction for young workers show just a minor effect, and sometimes even none at all, on occupation and wages for the targeted group. This implies that the windfall money does just in a fractional part accrue the workers in the treated age. Hence the question about what had happened to the windfall money was the big unresolved question the trio tried to solve. To do so the authors used an annual integrated data register at individual and firm level containing variables such as individual working time, wages, level of education, gender, time of birth and firm identifiers. For each firm, they also used a vast number of key variables such as margins, etc.

When investigating the effect on firm performance and employment effect, Saez, Schoefer, and Seim divide all firms into four quantiles based on the proportion of total wage-earning paid to young workers aged 19-25 in 2006. Of the four quantiles, the lowest contains a vast spread of share of young employees and is hence not suitable for comparison. The middle two quantiles are combined into the control group “medium share young employees” and the top quantile is used as treatment group “high share young employees”, sometimes divided into two groups “fairly high share” and “very high share” respectively. The comparison of the performance is then made using a conventional DiD-model on balanced panels for the years 2003-2013. The data is however converted to normalized values using 2006 as base year.
Using the same definition of employment as previous, i.e. a lower income threshold, the authors find significant employment effects of the reform. Employment in all ages increases within firms with a “high share of young workers” at 4.6 percent and more than twice as much for “very high share” than “fairly high share”, all in comparison with the “medium share of young employees”. The evaluation of firm performance effects shows similar estimates, were for example profit (EBIT) shows a significant increase of 8 percent in benchmark. These result point to that some of the windfall money is used for new hiring, among all layers of age, and some are used for investment or capital accumulation for the firms and its owners. This conclusion is shared with previous studies.

However, when Saez, Schoefer, and Seim evaluate the average wage effects within firms, this conclusion gets more ambiguous. Observe that the average wage effects within firms weren’t evaluated in the studies of Eggebark and Kaunitz (2017) or Skedinger (2014), who just evaluated the wage effect for youths. The change of dependent variable is motivated by the hypothesis that, due to Sweden’s quite rigid and norm-based wage setting system, it is more likely that there is a general rent sharing effect on wage increases. Rather then that the treated workers receive the whole wage increase. The results strongly support this hypothesis. Estimates show a gross increase of average wages within firms in the magnitude of the full amount of windfall money. This suggests that reform generates a net surplus of wealth. Such effects might seem like “pulling a rabbit out of a hat”, but it may have reasonable explanations as for instance credit restrictions. When firms receive windfall money, it can be used for more credits.

The most recently published evaluation of the payroll tax reduction for youths in Sweden was made by Daunfeldt, Gidehag, and Rudholm (2018). They argue that the effect of the reform should not be estimated upon relative treatment intensity among firms, but instead by the actual nominal treatment intensity. By using the definition made by Saez, Schoefer, and Seim a firm with just one young employee can get higher treatment intensity than a big firm with twenty-one. This is according to Daunfeldt, Gidehag, and Rudholm creating a bias in the calculations on how many jobs the reform created. In essence, the methodology made by Daunfeldt, Gidehag, and Rudholm (2018) is the same as in Saez, Schoefer, and Seim (2017). However, when the later studies the effect of the reform upon a wide range of independent variables, the authors of this article just study the employment effect in general, as well as for the targeted age group.
To overcome the suggested problem with previous studies, the authors divide all firms within the dataset that gained monetarily from the reform into five equally sized quintiles based on the nominal monetary gain, i.e. first quintile for the 20 percentage with least gain, second quintile contains the 20-40 percentage who gained second least and so on. The firms who had no employees between the age of 19-25 in 2006 are used as control group. The comparison was then made between the control group and all quantiles of firms treated firms separately. Here one thing is worth noticing: to divide comparison groups in this way may create bias due to big differences in other things than just the number of young employees that may separate the firms. For example, the number of young employees is highly correlated with firm size, which in turn is correlated with firm nominal expansion capacity. This bias should according to the authors be corrected by the use of a difference in difference in difference (DDD) model which takes previous employee development disparities between comparison groups to count.

The results of the estimations show a significant general employment effect for the three groups with the highest treatment intensity. The effects increase as treatment intensity increases, e.g. the third quintile which in median had a cost reduction of 25 000 kr per year hired 0.15 more employees and the fifth quintile which in median had a cost reduction of 101 000 kr per year hired approximately one more employee. The main body of the recruitment was of people in the targeted age range of 19-24. In summary, the authors argue that the reform generated approximately 16 400 new jobs, which are twice as many as Egebark and Kauntiz (2017) suggested. However, Egebark and Kauntiz were only focusing on new jobs for employees in the age of 19-24. Daunfeldt, Gidehag, and Rudholm estimate the number of new jobs, generated by reform, for workers in this age range to be approximately 13 000. It is also important to stress that the new jobs created, are due to the author’s definition of a job, not full-time employments. Although this is the case for all studies described in this chapter, who all have defined a job or being employed as a part-time job and part-time employment respectively.
5 Empirical approach

Overnight the reform created a vast drop in the hiring cost of young labor for the firms. This creates the foundation for researchers to execute a quasi-experiment. This has been exploited by previous evaluation of the reform, which has all been relying on a difference in difference (DiD) framework. In core, the DiD-methodology uses the evolution of a treatment group and a control group (Egebark and Kaunitz, 2017). In this specific case the desirable treatment group is all payroll-tax-subsidy-entitled young individuals. Correspondingly the worthwhile control group is a group of individuals with, so few disparities in their characteristics with the treatment group as possible, except the fact that they were not affected by the reform. If the two groups share the trend in the pre-reform period (in literature referred to as a parallel trend) it is assumed that, in absence of an intervention, they would do so also in the continuation of time. Based on this assumption, one can estimate whether the reform (in literature often referred to as treatment) had an impact on the variable of interest by comparing the two group’s evolution after the reform. This is done by examining the sign, value, and significance of the DiD-estimator ($\hat{\delta}_{DD}$).

$$\hat{\delta}_{DD} = (\bar{y}_{t=2}^{TG} - \bar{y}_{t=1}^{TG}) - (\bar{y}_{t=2}^{CG} - \bar{y}_{t=1}^{CG}) = \Delta \bar{y}^{TG} - \Delta \bar{y}^{CG} = (\theta^{TG} - \theta^{CG}) + (\varphi^{TG} - \varphi^{CG})$$

In the equation above the conventional definition of the DiD-estimator is on the left-hand side. The sample mean of the observed values of the variable of interest for the pre-period is denoted with $t = 1$ and the treatment period correspondingly $t = 2$. The treatment group is denoted with $TG$ and comparison group with $CG$. On the right-hand side of equation (3) is a definition of importance. Since the trend is defined by $\theta^i$ and the treatment effect is defined as $\varphi^i$ the intention is that $\theta^{TG} - \theta^{CG} = 0$, i.e. that the two groups share the same trend. When estimating the treatment effect of the treatment group relative to the control group, with a DiD-model, it is imperative that there exists no spillover effect, i.e. $\varphi^{CG} = 0$. If these conditions aren’t fulfilled the DiD-estimator will be biased and $\hat{\delta}_{DD} \neq \varphi^{TG}$ (Stock and Watson, 2011). It should be noted that it is not a necessary condition that the two groups share the same nominal values of the variable of interest $\bar{y}^i$, this can be corrected by group dummies. The reader is directed to the appendix AT1. to see the derivation leading to the DiD-estimator.
5.1 Common trend assumption and public data

There are, from what the public data offers, two obvious control groups. Firstly, the individuals in Sweden who weren’t affected by the reform, i.e. those above the age threshold. Secondly, no such reform was introduced in any other European country during the same period, which gives another possible comparison group, the youth living in other European countries. The previous discussion suggested that the DiD-estimator will be biased if either of the parallel trend assumption or the assumption about no spillover effects isn’t fulfilled. There is, in theory, a trade-off in choosing comparison group. The comparison group most likely to be homogenous to the treatment group is the Swedish individuals just above the age threshold, i.e. Swedish inhabitants at the age of 26\(^9\). However, a more homogenous, in relation to the treatment group, comparison group may suffer from a negative spillover effect since their relative price of labor has increased in relation to the treatment group. This could lead to an underestimation of the treatment effect.

**Figure 7 The treated group and publicly available comparison groups**

Y-axis describes the employment rate. In Figure 7 the three vertical lines describe in order from left: 1: The introduction of the reform. 2: The extension of the reform. 3: The repeal of the reform. * EU 20-24 is the European countries used as control group in tests for the treatment effect of the introduction. Source: Eurostat.

\(^9\) 27 after reform extension in 2009.
However, since the public data from SCB and Eurostat is aggregated and doesn’t contain single age cohorts, but instead five- and ten-year cohorts, this kind of luxurious trade-off problem is not assumed to be a topic of interest in this study. As figure 7 clearly describes, the problem lies instead in fulfilling the parallel trend assumption. It is in theory possible to overcome the problem with a common trend by adding underlying covariates that describe the disparities in trends, or by using some sort of matching procedure (Blundell et al., 2004). More complex matching methods such as propensity score matching, is feasible with a rigorous individual dataset, however not if the data is aggregated. Adding underlying covariates may seem like an intuitively appealing solution to this problem. However, it comes with several disadvantages.

It is hard to find variables that affect the employment or unemployment and ones that can be assumed to do so are in the public data frequently missing for numerous of the panel members, independent on which publicly available dataset used\(^\text{10}\). An even more severe problem is the endogeneity which can occur as a result of adding these variables. One of the key assumptions in the least square methodology is the exogeneity assumption, simply said this means that the error term should be uncorrelated with all of the explanatory variables \(\text{Cov}(x_{itj}, u_{itj}) = 0\), although it can be presented in other ways as well (Verbeek, 2004). This assumption is violated if any of the explanatory variables in the model are endogenous. When using covariates that have been proved to correlate with employment/unemployment, like for example GDP growth (Reyenga and Bentolila, 1995), it is likely that the problem of simultaneity occurs. GDP growth is determined by the production in the country, which truly can be said to depend on the number of occupied workers in the country. Hence, the dependent and explanatory variables are simultaneously a function of each other, i.e. the GDP growth is an endogenous variable (Wooldridge, 2013). There are ways to overcome this problem. Conventional ways are to find exogenous instrument variables that correlate with the explanatory variable, also known as IV regression, or using some sort of simultaneous equations model, often referred to as SEM (Wooldridge, 2013). This is however a time-consuming procedure which due to the time limitations of this study haven’t been conducted.

The described circumstances create difficulties in finding control groups which fulfilled the parallel trend assumption. However, this problem is attempted to be handled in such a transparent way as possible by using two methodologic approaches, applied to the same data. The first is a normal least square DiD-approach, were the cardinal methodologic feature in

\(^{10}\) This is mainly a problem for the first pre-period i.e. 2001-2004.
order to achieve more accurate estimators, is to expand the dataset in longitude direction, i.e. using a panel dataset, and utilize the existence of unobserved effects by panel data methods. The broad bibliotheca of publicly available data on the field of labor market includes a vast amount of regional subdivisions of countries which makes this increase of observations possible.

The second approach to answering the research question is to conduct a synthetic control method. This is a newly established matching method that suits the evaluation of treatment effects on aggregated units\textsuperscript{11} (Abadie, Diamond and Hainmueller, 2011). The method originates from the idea that some combination of the control units can generate a synthetic control unit that incorporates the same characteristics as the treatment unit (Abadie, Diamond and Hainmueller, 2007). This is utilized by the synthetic control method by weighting the control units to create a synthetic control unit which approximates the relevant characteristics of the treated group in the pre-treatment period (Abadie, Diamond and Hainmueller, 2011). Since the two approaches differ considerably, similar results would imply that the estimators of the treatment effect are true. If any of the two methods fail to satisfy the necessary conditions for unbiasedness of estimator, then the other might be able to perform an unbiased estimate. In the following chapters, both methodologies will be further presented.

2.1 Data

In this study, three different balanced panel datasets for the years 2001-2008 and 2010-2016\textsuperscript{12} have been used. They are firstly the Swedish counties dataset, secondly the European countries dataset and thirdly the European region dataset. The ambition has been to test the treatment group against both the two possible comparison groups. Where the first dataset is used to test the treatment group against their slightly older compatriots, the second and third datasets are used to test the treated Swedish youth against other Europeans in the same age cohorts. These two datasets complement each other were the former is further stretched in the time perspective whereas the later is further stretched in the longitude direction. Each of the dataset’s origin from different data sources. The Swedish counties dataset origins from the annual reported labor

\textsuperscript{11} Observe the denotation here, a unit is just one single panel member, whereas a group is a bunch of panel members, i.e. a group of units. Due to methodology differences groups are often discussed when describing the linear least square DiD approach, while units are often discussed when describing the Synthetic control approach.

\textsuperscript{12} It should be stressed that the panel members in the European Countries and European Regions datasets marginally differs between the two time periods and depending on which dependent variable is tested.
statistics based on administrative sources (RAMS)\textsuperscript{13}. The European countries dataset is collected from the quarterly reported Labour force survey (LFS), the European region dataset is based on the annual reported Regional labor market statistics. The two last are distributed by Eurostat, whereas the first is published by SCB. So that no confusion may occur in reading the estimates the three datasets will henceforth just be denoted RAMS-, Countries-, and Regions-dataset.

Table 1 Panel observation summary

<table>
<thead>
<tr>
<th>Countries\textsubscript{emp}</th>
<th>Countries\textsubscript{unemp}</th>
<th>Regions\textsubscript{emp}</th>
<th>Regions\textsubscript{unemp}</th>
<th>RAMS\textsubscript{emp}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>21</td>
<td>17</td>
<td>243</td>
<td>157</td>
</tr>
<tr>
<td>$n_2$</td>
<td>29</td>
<td>23</td>
<td>248</td>
<td>239</td>
</tr>
<tr>
<td>$T$</td>
<td>32</td>
<td>32</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

In Table 1 $n_1$ denotes the observed amount of panel members for the tests of the introduction, i.e. 2001-2008. $n_2$ denotes ditto but for the tests of the repeal period, i.e. 2010-2017. $T$ simply denotes the time periods which each dataset contains for the tests.

The LFS is a survey-based dataset were a sample of the population of working age (15-74) in each reporting country is collected. The data collection is made by each country’s national bureau of statistics and the samples are continuously resampled. The respondents are asked questions about their labor force participation during a reference week. All countries use the same standardized methodology. Numerous variables are collected and averages over different time spectra as well as for specific subgroups can be calculated. For a further description of the data collection, the reader is remitted to either the LFS method and definition paper (Eurostat, 2001) or the individual bureau’s documentation such as SCB’s paper about statistical manufacturing of the LFS (SCB, 2019).

The RAMS is, as the name describes, based on data from Swedish administrative authorities, and the labor market data is mainly based on annual income. From this dataset, no evaluation of the unemployment rate is possible, and the definition of being employed differs from the LFS since it is based on annual income. Even though the income thresholds for being registered as employed are aimed to be set by SCB in such a way that it should be comparable with the LFS for all age cohorts (SCB, 2005), this target is only partly fulfilled. As an example, in 2007 is the LFS reporting the annual employment rate for Swedish inhabitants in age 20-24 to be 52 percent.

\textsuperscript{13} On Swedish "Registerbaserad arbetsmarknadsstatistik" hence the acronym RAMS.
The RAMS, on the other hand, is reporting the Swedish counties’ average employment rate\textsuperscript{14} among same age cohort to be 62 percent, where no county has an employment rate as low as 52 percent.

It should be made clear that the Regional labor market statistics also origins from the LFS. The two data sources just differ in time and longitude spectra as well as the age cohorts. The publicly available LFS data (as well as the RAMS data) is reported in five years age cohorts\textsuperscript{15}, whereas the Regional labor market statistics are reported in teen years age cohorts. Hence, when using the Countries dataset, the treatment and control group will be in age cohort 20-24, whereas they are 15-24 in the Regions dataset. The country division in the Regional labor market statistics is based on the Nomenclature of Territorial Units for Statistics (NUTS)\textsuperscript{16} and goes down to the NUTS2 level. In order to acquire balanced panels, the problem with missing observations in the Regional labor market statistics has been handled by step by step reducing the level of the subdivision, so that the whole comparison group still is covered. When for example the German regions (NUTS2) Chemnitz and Leipzig have missing values, both these and the region Dresden are replaced with the German state Sachsen (NUTS1) which includes these three regions.

2.2 The least square approach

The interest of the estimations conducted in this study lies in estimating the partial effect of one of the observable explanatory variables in the population regression function $E(y \mid x_1, x_2, \ldots, x_k, c)$. In order to get an accurate estimate, it is desirable to keep all other factors that affect the dependent variable constant (Wooldridge, 2002). As earlier described the problem of endogeneity among suggestible explanatory variables due to simultaneity is severe. This makes the estimation of the partial effect of the treatment more problematic since it causes issues with including covariates\textsuperscript{17}. However, by using panel data one can, in general, obtain more efficient estimations since there exist methods that neatly can, at least to some extent, solve for the unobserved effect (Verbeek, 2004).

In general, the main reason for using panel data methods is to allow the explanatory variables to be correlated with the unobserved effect (Verbeek, 2004). This is feasible when violations of

\textsuperscript{14} When using the RAMS and Regions datasets the treatment group is not the Swedish average employment rate, but instead the individual counties or regions employment rate. Same applies for unemployment.
\textsuperscript{15} Consequently, in the RAMS dataset the treatment group is in age cohort 20-24, whereas control group is 30-34.
\textsuperscript{16} On French “Nomenclature des unités territoriales statistiques” hence the acronym NUTS.
\textsuperscript{17} This is at least the case if one uses any standard linear least square approaches.
the exogeneity assumption due to including a suggestable explanatory variable comes from omitted variable bias from time-constant factors. However, in the case which this thesis is studying, the suggestable explanatory variables such as differences in GDP growth between countries or regions in the dataset is not only correlating with some time-constant factor between the panel members, but are also correlating with other time inconstant factors.

Although the usage of panel data methods, in this specific case, does not support the usage of more explanatory variables, it comes with other advantages. As earlier stated, it is desirable to hold all other explanatory variables constant when obtaining the estimator of the partial effect. If we assume that some of the unobservable effect \( c \) is time-constant, but between the panel members deviating, then it is possible to hold this part of the unobserved effect constant

\[
E(y_t | x_t, c) = \beta_0 + \beta_0 x_t + c \quad \text{(Wooldridge, 2002)}
\]

This mechanism is already used in the simple DiD-model, where the group dummy corrects for the time-constant variation between the treatment and comparison group. However, the standard group dummy doesn’t control for time-constant unobserved effects between the units within the two groups.

On micro level individual data, standard economic textbooks often suggest the usage of treating the unobserved effect \( c \) as either a “fixed effect” or a “random effect”. Even though these two methods are the most discussed, “first difference” can also be used to handle the unobserved effects. When possible, it is preferable to treat the unobserved effect as a “random effect” since it generally is more efficient (Wooldridge, 2013). But to treat the unobserved effect as a “random effect” needs a more stringent set of assumptions to be fulfilled, in order to be consistent. It is commonly suggested that the doubtful researcher should conduct a Hausman test to choose how to treat the unobserved effect.

The data used in this thesis is not micro level individual data, it is aggregated to a country/regional level. Hence, one cannot treat this sample as a random sample from a large population (Wooldridge, 2013). Instead, every unit is “one of a kind”, this fact rules out the usage of a random effects approach since it relies on that the sample is random (Verbeek, 2004).

Another sometimes mentioned method is “first difference”. This method has some advantages, as well as pitfalls. If all necessary assumptions, later declared, are fulfilled both “fixed effects” and “first difference” are unbiased and consistent. Hence, standard econometric textbooks suggest that the choice of which estimation method to use lies in how efficient estimator it supplies. In the presence of autocorrelation, the “first difference” generally produces better estimates (Wooldridge, 2013). These arguments will hold if the number of observations that
were affected by the reform was sufficiently many, but for instance in the Countries-dataset Sweden is the only unit affected by the reform. If one assumes that the reform had a direct effect that lasted until the repeal and applied the standard setting of the first difference-model, this will just generate one observation of the effect (the first period of the treatment time for the treated unit). When using a fixed effects-model all observations of the treated unit during the treatment period will show the effect. Hence the fixed effects approach is chosen in this thesis. In this case, it seems superior to other standard panel data methods and some of the problems associated with autocorrelation can be handled with the use of robust standard errors. Fixed effects-models are in fact the most conventional model for policy evaluation, all articles in litterateur review use this approach and it is most often superior to other panel data methods for policy evaluation (Wooldrige, 2002).

Due to the possible issues discussed earlier, simultaneity in suggestable covariates, aggregation of data, etc. the fixed effects-model used in this thesis is of the most basic sort, containing no covariates. The explanatory variables are an indicator variable for the period, an interaction of the indicator variable for treatment period and treatment group and the “fixed effects”. The plain fixed effects-model used is presented below in equation (4).

\[ y_{tij} = \beta_1 P_t + \delta_{DD}(P_t \times G_t) + \alpha_j + \epsilon \]

The dependent variable \( y_{tij} \) has been either the employment rate in the population or the unemployment rate. The indices \( t, i \) and \( j \) stands for time, treatment group identification and country/region identification respectively. The period dummy \( P_t \) takes value 0 if the observation is in the pre-treatment period and 1 in the treatment period. Correspondingly the group dummy \( G_t \) takes value 0 if observation belongs to the control group and 1 if it belongs to the treatment group. The interaction dummy, effecting the DiD estimator takes thereby value 1 only if the observation is in treatment period and belongs to the treatment group. In the model the “fixed effects” denoted \( \alpha_j \) captures all time-constant variations in \( y_{tij} \) which differs over the countries or regions, indexed \( j \). Simpler said \( \alpha_j \) can be seen as each geographical units own independent intercept.
The fixed effects-model used in this thesis imposes that additional to the two already established assumptions (absence of spillover effect and the parallel trend assumption) more assumptions must be fulfilled to acquire unbiased and consistent estimates of the reform effect (Wooldridge, 2013).

Firstly, the model must be linear and contain an estimator of the unobserved effects, this is due to model-specification meet. Secondly, the sample from each geographical unit must be random, due to the data collection described in the previous subchapter this is assumed to be fulfilled. Thirdly, there must exist no perfect multicollinearity and each explanatory variable must change over time (for at least some j) this assumption is due to model-specification fulfilled. This condition is the reason why neither an intercept or a dummy indicating treatment, is included in the model, both are taken care of by the “fixed effects”. Fourthly the strict exogeneity assumption $E(\epsilon_{it}|X_i, \alpha_i) = 0$ must be fulfilled, a crucial assumption for all least square’s estimations. However, is it a tricky assumption to satisfy. It can occur from for example measurement errors, omitted variables bias or simultaneity (Wooldridge, 2013). The problem of simultaneity has previously been discussed and there may be a trade-off between the omitted variable bias and the simultaneity bias. Neither of the problems is appealing. Additionally, the fact that the panel is balanced makes interpretation more straightforward, even though this is not a necessary condition.

However, if these four assumptions are met, as well as the two DiD specific assumptions, then the estimator given by the fixed effects-model is both unbiased and consistent. By fulfilling two additional assumptions the estimator is not only effective but also fulfilling all the Gauss-Markov criteria making it the best linear unbiased estimator (BLUE). These two assumptions are firstly a homoscedastic distribution of the error term $Var(\epsilon_{it}|X_i, \alpha_i) = Var(\epsilon_{it}) = \sigma^2_{\epsilon}$ for all $t = 1, \ldots, T$. Simply, a constant distribution of error term. The second of these assumptions is the absence of autocorrelation $Cov(\epsilon_{it}, \epsilon_{it}|X_i, \alpha_i) = 0$ for all $t \neq t´$ which in other words means no serial correlation in the error term (Wooldridge, 2013).
It is not likely to assume that the simple-DiD model used in this thesis perfectly predicts the outcome of the dependent variable in every single country or region, for all time periods. Even though the time average prediction error for every single country or region can be assumed to be zero due to the fixed effects-model, it cannot be assumed that the model catches individual, for countries or regions, systematic deviation of the dependent variable, due to variation in the underlying unobserved variables. This may, for example, be some country-specific shock. Systematically underprediction (or overprediction) yields that the error term $\varepsilon_{it}$ is correlated over time $\text{Cov}(\varepsilon_{it}, \varepsilon_{it}' | X_i, \alpha_i) \neq 0$ for a given country or region. Observe that even if this violation of the assumption exists, the estimator is still unbiased, given that the first four assumptions were fulfilled. The problem is instead a matter of inference. Fortunately, by using cluster-robust standard errors one can achieve inference which is robust to arbitrary heteroskedasticity and autocorrelation. The reason to use cluster robust standard errors, where the clusters are each comparison unit, instead of robust standard errors is simply that it is likely that the errors, for example, Sweden 2004 correlates with errors for Sweden 2003 rather than Ireland 2003, i.e. the autocorrelation in the error term is within the countries or regions (Cameron, Miller, 2015).

The two DiD specific assumptions in order to get an accurate DiD-estimator is firstly the absence of spillover effect $\varphi^{CG} = 0$. This is a rather tricky assumption to test since the DiD-estimate is the relative difference between the comparison and treatment group. Here the researcher must rely on intuition. However as earlier stated the spillover effect can, in this case, be assumed to be rather marginal. Firstly, swedes in age cohort 30-34 can’t be assumed to have the same set of characteristics as the ones in age 20-24. Secondly, other European youths can’t be assumed to be affected by the reform in Sweden.

The other DiD-assumption, i.e. the parallel trend assumption is somewhat easier to test. It can either be distinguished by observing the two trends visually or by conducting a placebo treatment. The placebo treatment is implemented to the dataset for the time period, or periods, just before the real treatment was set in place. In addition, the researcher eliminates all observations for the time periods as the real treatment was active. If the placebo effect is insignificant one can assume that the two comparison groups share a common pre-period trend, hence, fulfilling this necessary assumption.
2.3 The synthetic control group approach

With the ideal dataset, a single comparison unit or a sum of several units can generate an unbiased and consistent estimator of the treatment effect. However, when the ideal dataset isn’t supplied other estimations methods can provide better estimates. In a panel data model approach, the problem with a non-parallel trend in the pre-period could, for instance, be solved by choosing just those panel members which share a parallel trend in the pre-period. Such methods are not new but when the selection is done by the researcher’s own head it may create selection biases. Hence, even though it may seem intuitively right to compare Sweden only to other Nordic countries, to do so may result in selection bias. Additionally, a standard inference technique doesn’t take the uncertainty associated with the generation of the synthetic control group into account. Hence, choosing a control group based on which units are strongest correlated with treatment the unit or units in the pre-period tends to cause problems with interpretation of the significance of the result. Addressed problems have however Abedale and Gardeazabal (2001) attempted to solve. They constructed a data-driven selection methodology which also includes inference techniques taking the selection uncertainty into account. The methodology has been further developed and the publication of the Synth package\(^{18}\) by Abadie, Diamond, and Hainmueller (2011) and the Synth_runner package\(^{19}\) by Galiani and Quistorff (2016) made the synthetic control method disposable for all researcher.

By using a weighted combination of multiple control units from the donor pool, this method allows the researcher to construct a synthetic control group that can approximate the relevant characteristics of the treated group during the pre-treatment period. This is done by weighting the observed control units in such a way that the synthetic control group in the best possible way imitates the treatment group during the pre-period. The synthetic control unit is then used for comparisons with the treatment unit in the post-period.

Suppose that we observe units \(J = 1, 2, \ldots, J + 1\) for period \(t = 1, 2, \ldots, T\). We assume that only the first unit is exposed to the treatment. Consequently, the remaining units \(J = 2, 3, \ldots, J + 1\) are the donor pool, i.e. can contribute to the construction of the synthetic control unit. Further, we assume that the treatment occurs at the time period \(T_0 + 1\) so that \(1, 2, \ldots, T_0\) are the pre-periods and \(T_0 + 1, T_0 + 2, \ldots, T\) are the treatment periods. We define two possible outcomes. \(Y_{it}^N\) which refers to the outcome which would have been observed for unit \(i\) at time \(t\) in the

\(^{18}\) https://web.stanford.edu/~jhain/Synth

\(^{19}\) https://github.com/bquistorff/synth_runner
absence of treatment, whereas $Y_{it}^l$ refers to the outcome that would have been observed if unit $i$ at time $t$ is exposed for treatment. The target is to estimate the effect of the reform, i.e. the difference between these two outcomes for treatment periods $T_0 + 1, T_0 + 2, \ldots, T$. Since we assume that the treatment didn’t affect unit $i$ during the pre-period we assume $Y_{it}^l = Y_{it}^N$ as $t \leq T_0$. However, we do assume that the treatment affected unit $i$ during the treatment period, thus $Y_{it}^l = Y_{it}^N + \alpha_{it}$ when $t > T_0$. This implies that $\alpha_{it}$ is the parameter intended to be estimated. Since we are interested in the effect of the treatment on the treated unit, i.e. $\alpha_{1t}$, and its treatment period outcome is observed $Y_{i1t}$ for $t > T_0$ we only need to estimate $Y_{i1t}^N$ to estimate $\alpha_{1t}$ (Abadie, Diamond and Hainmueller, 2011).

In order to do so, we make use of the observed outcomes in pre-periods from the units which never were affected by the reform to construct a synthetic control group i.e. $Y_{it}^N$. We define a $(T_0 \times 1)$ vector $K = (k_1, k_2, \ldots, k_{T_0})'$ that denotes some linear set of pre-period outcomes such that $\bar{Y}_i^K = \sum_{s=1}^{T_0} k_s Y_{is}$. In order to control for time varying factors which are equal across the units, it is possible to include $M$ (where $M \leq T_0$) linearly independent combinations of pre-period outcomes\(^{20}\) such that $\bar{Y}_i^{K1} = \sum_{s=1}^{T_0} k_s Y_{is}^{K1}, \ldots, \bar{Y}_i^{KM} = \sum_{s=1}^{T_0} k_s Y_{is}^{KM}$. In order to clarify the meaning of this one can think of the simplest case where there is just one set of linear pre-period outcomes (and all pre-periods included) denoted as $K$. Then $k_1 = k_2 = \ldots = k_{T_0} = 1/T_0$, thus $\bar{Y}_i^K = 1/T_0 \sum_{s=1}^{T_0} Y_{is}$ is just the average of the outcomes for all units in all pre-periods (Abadie, Diamond and Hainmueller, 2007). It is also possible to include a vector of all observed covariates for each unit $U_i$, but due to already discussed reasons, this is not done in this thesis.

The next step in the construction of our synthetic control group is to define a $(J \times 1)$ vector of weights $W = (w_2, w_3, \ldots, w_{j+1})'$ such that $w_j \geq 0$ for $j = 2, 3, \ldots, J + 1$ and that $w_2 + w_3 + \ldots + w_{j+1} = 1$. Each $w_j$ represents one weighted average of one particular control unit. The fundamental idea is to choose such vector of weights $W^*$ so that the resulting synthetic control group best approximates the treatment group with respect to the outcome variable in the pre-treatment period. $W^* = w_2^* + w_3^* + \ldots + w_J^*$ is chosen so that $\bar{Y}_i^{K1} = \sum_{j=2}^{J+1} w_j^* Y_{ij}^{K1}$ or $\bar{Y}_i^{K1} = \sum_{j=2}^{J+1} w_j^* Y_{ij}^{K1}, \ldots, \bar{Y}_i^{KM} = \sum_{j=2}^{J+1} w_j^* Y_{ij}^{KM}$ if multiple vectors of pre-period outcomes are used (Abadie, Diamond and Hainmueller, 2011). If so is approximately the case, it yields that using:

\(^{20}\) In this thesis $M = T_0$ in order to fully control for all time varying factors which are equal across all units.
\( (5) \quad \hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{jt} \)

for periods \( T_0 + 1, T_0 + 2, \ldots, T \) as an estimator of \( \alpha_{1t} \) is appropriate (Abadie, Diamond and Hainmueller, 2007).

In this thesis, the previously described procedure has been conducted through the `synth_runner` package\(^{21}\) in the statistical software STATA. In the general case,\(^{22}\) the `synth_runner()` function finds the optimal vector of weights \( W^* \) by combining the characteristics of the treated unit in the \((k \times 1)\) matrix \( X_1 = (U'_1, \bar{Y}_1^{K1}, \ldots, \bar{Y}_1^{KM})' \) with the same characteristics of the control units in the \((k \times j)\) matrix \( X_0 \) which at the \( j - th \) row \((U'_j, \bar{Y}_j^{K1}, \ldots, \bar{Y}_j^{KM})' \). In practice `synth_runner()` function minimize:

\( (6) \quad ||X_1 - X_0 W||_V = \sqrt{(X_1 - X_0 W)'V(X_1 - X_0 W)} \)

Where the \( V \) is defined as a \((k \times 1)\) symmetric and positive semidefinite matrix. The reason why the \( V \)-matrix is introduced is to allow different weight to the variables in \( X_0 \) and \( X_1 \) depending on their prediction ability on the outcome. There may be occasions when the researcher has motives to reset the weights of his or her own. (Abadie, Diamond and Hainmueller, 2011). Such a procedure has however not been done in the work with this thesis. Instead, has the \( V \)-matrix been set in such a way that it minimizes the mean square predictive errors (RMSPE) of the outcome variable for the treated unit in the pre-periods through the data-driven default procedure in `synth_runner()`.

After creating a synthetic control unit. It may seem reasonable to estimate the treatment effect \( \alpha_{1t} \) and determine the statistical inference through a least square procedure. However, when doing so one doesn’t take the uncertainty associated with the generation of the synthetic control group into account. Instead the \( \hat{\alpha}_{1t} \) is obtained within the `synth_runner()` function, by the treatment effect estimator function \((5)\), whereas the statistical significance is determined by estimating the same model on each untreated unit \( j = 2, 3, \ldots, J + 1 \) assuming it was treated at

---

\(^{21}\) Even though the `synth` package also been used complementary since the `synth_runner` package doesn’t report which units from the donor pool that have been assigned with a weight exceeding zero.

\(^{22}\) Observe that this is not the general case since no covariates is used.
the same time. This will produce a distribution of longitude-placebo treatment effects. This is also done by default by the \texttt{synth_runner()} function. If the distribution of longitude-placebo treatment effects incorporates many longitude-placebo treatment effects as large or larger than the main estimate $\hat{\alpha}_{1t}$ it is likely that the estimated effect only was observed by coincidence. For $\hat{\alpha}_{1t}$ the distribution of corresponding longitude-placebo treatment effects is $\hat{\alpha}^{PL}_{1t} = \{ \hat{\alpha}_{j,t} : j \neq 1 \}$. The two-sided p-values are then:

\begin{equation}
(7) \quad p-value = Pr( |\hat{\alpha}^{PL}_{1t}| \geq |\hat{\alpha}_{1t}| ) = \sum_{j \neq 1} \frac{1(|\hat{\alpha}_{j,t}| \geq |\hat{\alpha}_{1t}|)}{j}
\end{equation}

Inference application in the synthetic control method also supports looking at the joint significance of the treatment effect for all included treatment periods. This is done by studying the distribution of the $RMSPE_{\text{treat}}/RMSPE_{\text{pre}}$. between the different units. If the treatment unit’s $RMSPE_{\text{treat}}/RMSPE_{\text{pre}}.$ – value is in the tail of the distribution it implies a significant treatment effect (Abadie, Diamond and Hainmueller, 2007). This kind of test is however only useful if the estimated treatment effect consistently showing the same sign, and consequently, the time-average estimated treatment effect deviates from zero.

Additionally, the synthetic control method allows the researcher to conduct a time-placebo test. This is simply done in a similar way as a placebo test in a conventional least square by conducting the test on a time period before the true treatment was set in place. Since one must assume that the treatment effect occurs first after the treatment is set in place, the estimated time-placebo treatment effect should be zero, i.e. $\hat{\alpha}_{1t} = 0$ for $t \leq T_0$. If this is not even approximately the case one must assume that the synthetic control method generates an unprecise synthetic control unit.
6 Estimation and Results

The methods described in the previous chapter are used on all feasible datasets to estimate the effect on employment rate and unemployment rate of both the introduction and repeal of the reform. Observe that the Synthetic control method could not be used on the RAMS dataset since the synthetic control method relies on that the characteristics of the treated unit are to find in the donor pool. When all units in the donor pool are quite homogenous to one another but similarly heterogeneous to the treatment units the data doesn’t support the usage of the synthetic control method. Neither could any unemployment test be done upon the RAMS dataset since it only measures employment.

The two treatment periods have in general\textsuperscript{23} been defined as the two first years after the intervention, i.e. the introduction or repeal of the reform. The pre-periods, which all treatment effects are relative to, are in the conventional fixed effects model defined as 2001-2004 for the test of the introduction effect and correspondingly 2010-2013 for the repeal effect. This so that placebo tests for the two pre-period years directly before the intervention could be conducted without changing the pre-period. When estimating the treatment effect through the Synthetic control method the full pre-period 2001-2006\textsuperscript{24} and 2010-2015 respectively, are used for the generation of the synthetic control unit.

The treatment group differs between tests conducted, not only on different datasets but also depending on which method is used. The conventional fixed effects model supports multiple treatment units. This allows the treatment group in tests conducted on the Region dataset to be eight independent Swedish regions and on the RAMS dataset to be 21 independent Swedish counties. The synthetic control method does however not support multiple treatment units wherefore the treatment unit applied on the Region dataset are reduced to just be the Swedish annual average.

\textsuperscript{23} This is however not the case for the test of the treatment effect of the introduction conducted on the Countries dataset. Since it may be consistent quarterly differences in examined variable outcome the pre- and treatment periods as been set in such way that it should not be any unbalances in amount of quarters in the two periods. As an example, the test of treatment effect in 2007 contains no observations from the two first quarters.

\textsuperscript{24} For the test of the treatment effect of the introduction conducted on the Countries dataset the pre-period stretches until second quarter 2007.
6.1 Results from the fixed effects model

None of the results generated from the tests conducted through the fixed effects model provides any significant results\footnote{On a 5\% significance level.} which also satisfies the parallel trend assumption, i.e. insignificant placebo tests. The estimated treatment effects on the employment rate of the introduction are presented in Table 2. One can see that the tests conducted on the Countries and RAMS datasets for the introduction of the reform provides estimated treatment effects are in line with expectations, though not significant or not meeting the parallel trend assumption, whereas the test conducted on the Region dataset provides ambiguous results. This is also visualized in Figure A1-A6, in appendix. Observe that all reported estimates are in percentage points.

<table>
<thead>
<tr>
<th>Period</th>
<th>Countries</th>
<th>Regions</th>
<th>RAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/06</td>
<td>-0.199</td>
<td>-1.26</td>
<td>-0.3609</td>
</tr>
<tr>
<td>05</td>
<td>-0.368</td>
<td>-2.073</td>
<td>-1.822***</td>
</tr>
<tr>
<td>06</td>
<td>-0.031</td>
<td>2.828**</td>
<td>1.55 ***</td>
</tr>
<tr>
<td>07</td>
<td>2.502*</td>
<td>1.026</td>
<td>3.445***</td>
</tr>
<tr>
<td>08</td>
<td>2.119</td>
<td>-0.068</td>
<td>1.438**</td>
</tr>
<tr>
<td>07/08</td>
<td>2.413*</td>
<td>0.479</td>
<td>2.441***</td>
</tr>
<tr>
<td>$N^1$</td>
<td>210</td>
<td>1215</td>
<td>210</td>
</tr>
<tr>
<td>$N^2$</td>
<td>420</td>
<td>1215</td>
<td>210</td>
</tr>
<tr>
<td>$N^3$</td>
<td>504</td>
<td>1458</td>
<td>252</td>
</tr>
</tbody>
</table>

p < 0.001***, p <0.01**, p<0.05*, p<0.1’

Table 3 presents the estimated treatment effect on the employment rate from the repeal period. The tests conducted on the Region and RAMS dataset fails to satisfy the parallel trend assumption. Although the test conducted on the Countries dataset satisfies the parallel trend assumption, it does not provide significant estimates, nor do any of the tests provide estimates in the expected direction. The results are however not that surprising when remembering that the control group is not as homogenous to the treatment group as one would desire. This was earlier visualized by Figure 7 and is as earlier described the main reason why the fixed effects model is complemented by the synthetic control method.
### Table 3 Employment rate (repeal)

<table>
<thead>
<tr>
<th>Period</th>
<th>Countries</th>
<th>Regions</th>
<th>RAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>14/15²</td>
<td>0.696</td>
<td>3,931***</td>
<td>3,242***</td>
</tr>
<tr>
<td>14¹</td>
<td>0.992</td>
<td>3,782***</td>
<td>2,187***</td>
</tr>
<tr>
<td>15¹</td>
<td>0.399</td>
<td>4,081***</td>
<td>4,297***</td>
</tr>
<tr>
<td>16¹</td>
<td>1,348</td>
<td>4,212***</td>
<td>6,833***</td>
</tr>
<tr>
<td>17¹</td>
<td>0,181</td>
<td>3,856***</td>
<td>8,525***</td>
</tr>
<tr>
<td>16/17²</td>
<td>0,763</td>
<td>4,034***</td>
<td>7,768***</td>
</tr>
<tr>
<td>N¹</td>
<td>580</td>
<td>1240</td>
<td>210</td>
</tr>
<tr>
<td>N²</td>
<td>696</td>
<td>1488</td>
<td>252</td>
</tr>
</tbody>
</table>

\[ p < 0.001^{***}, p < 0.01^{**}, p < 0.05^{*}, p < 0.1’ \]

In Table A1 and Table A2 in the appendix, the estimates of the treatment effect of unemployment are presented. Neither of them provides significant results that satisfy the parallel trend assumption. It may seem strange that the estimated treatment effect on the employment rate and unemployment rate can differ that much. However, it should be stressed that the unemployment rate depends on the size of the labor force and is not the inverse of the employment rate.

#### 6.2 Results from the synthetic control method

The estimated treatment effects on employment from the synthetic control method are rather small and aren’t consistently showing the expected signs. This is clearly visualized in Figure 8-13. The first three figures show the results from the tests on the treatment effect on the employment rate at the introduction of the reform on the Countries dataset whereas the three following presents the results at the same periods and variables on the Region dataset. In appendix, the same figures for the repeal period are presented in figure A7 – A12. Figures 10 and 13 describe both clearly that the estimated treatment effects on employment rate of the introduction are quite small. They also describe that the estimated longitude placebo effects are in some cases much bigger for other countries or regions.
Observe that the value on the Y-axis for figure 8, 9, 11, 12 does all describes the employment rate whereas figure 10 and 13 the value on the Y-axis describe the estimated treatment effect in percentage points.
Figures 9 and 12 visualize the time-placebo test when employment is the dependent variable and the tests are conducted on Countries and Regions dataset respectively. In the test conducted on the Countries dataset the placebo-synthetic control unit seems, as expected, to follow the observed employment rate for the placebo-treatment time. This suggests that in this case the synthetic control unit is precise. However, when looking at the same test applied to the Regions dataset the same conclusion can’t be made. In this case, the placebo-synthetic control unit deviates substantially from the observed values. This should be taken into concern when interpreting the estimates as an indication that the estimates probably is biased.

The result of the treatment effect on unemployment is not presented in any figure. It is presented in Table A5 and A6 though where one can clearly see that the time-placebo test shows considerable deviations between the placebo-synthetic control unit and the observed values for the placebo-treatment period. This is the case irrespective of which dataset used. Recapitulate that during this period no intervention was made, and hence, an unbiased synthetic control unit should be approximately equal to the observed value.

As earlier described, it is possible to test the joint significance of the treatment effects of all different time periods. This is however only of interest if one can assume that the estimates are unbiased and if the time-average treatment effect deviates substantially from zero. However, only one treatment effect estimator can be assumed to be unbiased (the treatment effect estimator of the introduction on employment from the test conducted on the Countries dataset) and this does not deviate considerably from zero \( \hat{\alpha}_{1t} = -0.066 \) for \( t = all T > T_0 \). Consequently, in this thesis, no test for joint significance is done.
7 Discussion

The result presented in the previous chapter suggests that none of the presumed unbiased estimators of the treatment effect could show any significant treatment effect. In fact, most of the estimators can according to the tests be assumed to be biased. Can one draw any conclusions about the treatment effect of this reform from this thesis? The answer is both yes and no. Most of the estimates are either inconsistent or not fulfilling some of the necessary conditions in order to be unbiased due to the problems with finding homogenous control groups. Hence is it hardly compatible with common practice to argue that the thesis finds any evidence that the reform had any effect on the labor market situation for young adults in Sweden. But the result doesn’t support the opposite argumentation either since the treatment effect estimator in so many of the tests must be assumed to be biased. However, one could argue that if the treatment effect of the intervention was of considerable size the treatment effect would be visible even with this type of data. But since the aggregated treatment effect wasn’t substantial (neither this study nor the previous suggests so), the treatment effect wasn’t visible.

So, do these results impose that the publicly published aggregated (macro) data isn’t as good for policy evaluation as microdata? I would say “yes it does” but at which extent is still unclear. What clearly can be concluded is that applying methods described in the thesis on the publicly published aggregated data, also previously described, one can’t find any unbiased and consistent results. It can be methods for which this can be done, although it seems to me quite unrealistic. During the work of this thesis, various methods have been tested such as instrument variables, transformation of data, and DDD-models. Additionally, both different time dummies and the time periods, defining the pre-period, have been tested in many different combinations. Neither of these elaborations has generated any results with substantial deviation from the ones presented in this paper. Unluckily for all researchers with a strict monetary restriction in general, and for myself in specific, it seems like reforms with such small effects on aggregate, as this reform had, must be studied with the use of microdata.

There is another thing that might, though, play a part in the different results in this thesis compared to previous studies: namely the definition of employment. Since all previous studies use quite generous definitions of being employed, their data contains far more employed persons than the data used in this thesis. If the reform affected the employment of part-time workers more than full-time workers, this would make the estimates out of datasets in this thesis
smaller and plausible more uncertain. According to Egebark and Kaunitz (2017) this is the case and consequently, this can be assumed to influence the estimations.

To conduct policy evaluation is of great importance, but it does not always contribute with as much clarity as wished. The purpose of this thesis was to evaluate both the effect of the reform introduction and repeal, as well as if the usage of publicly available aggregated (macro) data suits policy evaluation of this kind. Previous studies have contributed with some clarity about the magnitude of the introduction of the reform. To this subject, this thesis doesn’t contribute to any game-changing implications. The effect of the repeal can be seen as still unexamined by any suitable evaluation. However, I hope that this thesis has provided some more wisdom in the usage of publicly published aggregated (macro) data for policy evaluation.
References


**Data source**


APPENDIX

Figure A1 Countries emp. (Intro)

Figure A2 Countries emp. (Repeal)

Figure A3 Regions emp. (Repeal)

Figure A4 Regions emp. (Repeal)

Figure A5 RAMS emp. (Intro)

Figure A6 RAMS emp. (Repeal)

27 Observe that on this page the value on the Y-axis describes the estimated treatment effect, this applies for all figures on page.
Observe that the value on the Y-axis for figure A7, A8, A10, A11 does all describes the employment rate whereas figure A9 and A12 the value on the Y-axis describe the estimated treatment effect.
### Table A1 Unemp. (Introduction)

<table>
<thead>
<tr>
<th>Period</th>
<th>Countries</th>
<th>Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/06³</td>
<td>7.333***</td>
<td>7.772***</td>
</tr>
<tr>
<td>05²</td>
<td>6.630***</td>
<td>7.439***</td>
</tr>
<tr>
<td>06²</td>
<td>8.036***</td>
<td>8.104***</td>
</tr>
<tr>
<td>07¹</td>
<td>5.905**</td>
<td>8.318***</td>
</tr>
<tr>
<td>08²</td>
<td>7.105***</td>
<td>9.144***</td>
</tr>
<tr>
<td>07/08³</td>
<td>7.146***</td>
<td>8.731***</td>
</tr>
</tbody>
</table>

|            | N¹        |         |
|            | 220       | 1195    |
|            | N²        | 480      | 1195    |
|            | N³        | 524      | 1434    |

p < 0.001***, p < 0.01**, p < 0.05*, p < 0.1’

### Table A2 Unemp. (Repeal)

<table>
<thead>
<tr>
<th>Period</th>
<th>Countries</th>
<th>Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>14/15²</td>
<td>-0.490</td>
<td>-2.967***</td>
</tr>
<tr>
<td>14¹</td>
<td>-0.445</td>
<td>-2.599**</td>
</tr>
<tr>
<td>15¹</td>
<td>-0.540</td>
<td>-3.336***</td>
</tr>
<tr>
<td>16¹</td>
<td>-0.678</td>
<td>-2.863**</td>
</tr>
<tr>
<td>17¹</td>
<td>0.920</td>
<td>-2.430**</td>
</tr>
<tr>
<td>16/17²</td>
<td>0.121</td>
<td>-2.647***</td>
</tr>
</tbody>
</table>

|            | N¹        |         |
|            | 480       | 1195    |
|            | N²        | 524      | 1434    |

p < 0.001***, p < 0.01**, p < 0.05*, p < 0.1’
### Table A3 Employment rate (Countries dataset)

<table>
<thead>
<tr>
<th>Period</th>
<th>Treatment effect ($\hat{\alpha}$)</th>
<th>Period</th>
<th>Treatment effect ($\hat{\alpha}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>07$q_1$</td>
<td>-</td>
<td>16$q_1$</td>
<td>0.922</td>
</tr>
<tr>
<td>07$q_2$</td>
<td>-</td>
<td>16$q_2$</td>
<td>0.532</td>
</tr>
<tr>
<td>07$q_3$</td>
<td>-0.34</td>
<td>16$q_3$</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>(0.857)</td>
<td></td>
<td>(0.955)</td>
</tr>
<tr>
<td>07$q_4$</td>
<td>-0.076</td>
<td>16$q_4$</td>
<td>-0.047</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>08$q_1$</td>
<td>0.432</td>
<td>17$q_1$</td>
<td>0.508</td>
</tr>
<tr>
<td></td>
<td>(0.905)</td>
<td></td>
<td>(0.864)</td>
</tr>
<tr>
<td>08$q_2$</td>
<td>-1.312</td>
<td>17$q_2$</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>08$q_3$</td>
<td>1.81</td>
<td>17$q_3$</td>
<td>1.854</td>
</tr>
<tr>
<td></td>
<td>(0.523)</td>
<td></td>
<td>(0.773)</td>
</tr>
<tr>
<td>08$q_4$</td>
<td>-0.912</td>
<td>17$q_4$</td>
<td>1.521</td>
</tr>
<tr>
<td></td>
<td>(0.81)</td>
<td></td>
<td>(0.773)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>704</th>
<th>N</th>
<th>736</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RMSPE_{pre}$</td>
<td>0.922</td>
<td>$RMSPE_{pre}$</td>
<td>1.268</td>
</tr>
<tr>
<td>$RMSPE_{treat}$</td>
<td>1.137</td>
<td>$RMSPE_{treat}$</td>
<td>1.039</td>
</tr>
</tbody>
</table>

$\hat{\alpha}_{\text{placebo}}$ = average estimated placebo treatment effect under pre-period.
$\hat{\alpha}$ = average estimated treatment effect under treatment period.

*P*-value with a significance level of 5% in brackets, below estimates.
### Table A4 Employment rate (Regions dataset)

<table>
<thead>
<tr>
<th>Period</th>
<th>Treatment effect ($\bar{\propto}$)</th>
<th>Period</th>
<th>Treatment effect ($\bar{\propto}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>07</td>
<td>-0.263</td>
<td>16</td>
<td>-0.166</td>
</tr>
<tr>
<td></td>
<td>(0.905)</td>
<td></td>
<td>(0.938)</td>
</tr>
<tr>
<td>08</td>
<td>-0.7</td>
<td>17</td>
<td>-0.671</td>
</tr>
<tr>
<td></td>
<td>(0.816)</td>
<td></td>
<td>(0.816)</td>
</tr>
</tbody>
</table>

| $N$    | 1272                               | $N$    | 792                                |

<table>
<thead>
<tr>
<th>$RMSPE_{pre}$</th>
<th>0.000</th>
<th>$RMSPE_{pre}$</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RMSPE_{treat}$</td>
<td>0.748</td>
<td>$RMSPE_{treat}$</td>
<td>0.692</td>
</tr>
</tbody>
</table>

| $\bar{\propto}_{placebo}$ | 0.721    | $\bar{\propto}_{placebo}$ | 0.403    |
| $\bar{\propto}$           | -0.482   | $\bar{\propto}$           | -0.419   |

*P-value with a significance level of 5% in brackets, below estimates.*

$\bar{\propto}_{placebo}$ = average estimated placebo treatment effect under pre-period.

$\bar{\propto}$ = average estimated treatment effect under treatment period.

### Table A5 Unemployment rate (Regions dataset)

<table>
<thead>
<tr>
<th>Period</th>
<th>Treatment effect ($\bar{\propto}$)</th>
<th>Period</th>
<th>Treatment effect ($\bar{\propto}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>07</td>
<td>-2.518</td>
<td>16</td>
<td>-0.304</td>
</tr>
<tr>
<td></td>
<td>0.301</td>
<td></td>
<td>0.899</td>
</tr>
<tr>
<td>08</td>
<td>-2.382</td>
<td>17</td>
<td>-0.568</td>
</tr>
<tr>
<td></td>
<td>0.441</td>
<td></td>
<td>0.814</td>
</tr>
</tbody>
</table>

| $N$    | 752                                | $N$    | 1040                               |

<table>
<thead>
<tr>
<th>$RMSPE_{pre}$</th>
<th>0.117</th>
<th>$RMSPE_{pre}$</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RMSPE_{treat}$</td>
<td>3.467</td>
<td>$RMSPE_{treat}$</td>
<td>0.644</td>
</tr>
</tbody>
</table>

| $\bar{\propto}_{placebo}$ | 4.244    | $\bar{\propto}_{placebo}$ | 2.2469 |
| $\bar{\propto}$           | -2.45    | $\bar{\propto}$           | -0.436 |

*P-value with a significance level of 5% in brackets, below estimates.*

$\bar{\propto}_{placebo}$ = average estimated placebo treatment effect under pre-period.

$\bar{\propto}$ = average estimated treatment effect under treatment period.

*Observe that the unemployment estimates aren’t presented in figures.*
<table>
<thead>
<tr>
<th>Period</th>
<th>Treatment effect ((\hat{\alpha}))</th>
<th>Period</th>
<th>Treatment effect ((\hat{\alpha}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>07(Q_1)</td>
<td>-</td>
<td>16(Q_1)</td>
<td>-0.816 (0.816)</td>
</tr>
<tr>
<td>07(Q_2)</td>
<td>-</td>
<td>16(Q_2)</td>
<td>-0.028 (1)</td>
</tr>
<tr>
<td>07(Q_3)</td>
<td>-2.955 (0.19)</td>
<td>16(Q_3)</td>
<td>-0.924 (0.773)</td>
</tr>
<tr>
<td>07(Q_4)</td>
<td>-2.321 (0.381)</td>
<td>16(Q_4)</td>
<td>-1.483 (0.5)</td>
</tr>
<tr>
<td>08(Q_1)</td>
<td>-1.307 (0.571)</td>
<td>17(Q_1)</td>
<td>-0.007 (1)</td>
</tr>
<tr>
<td>08(Q_2)</td>
<td>-0.761 (0.714)</td>
<td>17(Q_2)</td>
<td>-0.425 (0.818)</td>
</tr>
<tr>
<td>08(Q_3)</td>
<td>-4.458 (0.333)</td>
<td>17(Q_3)</td>
<td>-0.283 (1)</td>
</tr>
<tr>
<td>08(Q_4)</td>
<td>-0.944 (0.762)</td>
<td>17(Q_4)</td>
<td>-1.22 (0.773)</td>
</tr>
</tbody>
</table>

\(N\) 704 \(N\) 736

\(RMSPE_{pre}\) 1.714 \(RMSPE_{pre}\) 1.177
\(RMSPE_{treat}\) 2.726 \(RMSPE_{treat}\) 0.884

\(\hat{\alpha}_{placebo}\) 4.365 \(\hat{\alpha}_{placebo}\) -2.644
\(\bar{\alpha}\) -2.124 \(\bar{\alpha}\) -0.648

*P*-value with a significance level of 5% in brackets, below estimates. 
\(\hat{\alpha}_{placebo}\) = average estimated placebo treatment effect under pre-period. 
\(\bar{\alpha}\) = average estimated treatment effect under treatment period. 
Observe that the unemployment estimates isn’t presented in figures.
### Table A7 Weights Employment rate (Countries dataset)

<table>
<thead>
<tr>
<th>Country</th>
<th>( w_j )</th>
<th>Country</th>
<th>( w_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czechia</td>
<td>0.201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>0.081</td>
<td>Ireland</td>
<td>0.07</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.022</td>
<td>Lithuania</td>
<td>0.066</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.088</td>
<td>Finland</td>
<td>0.542</td>
</tr>
<tr>
<td>Finland</td>
<td>0.608</td>
<td>UK</td>
<td>0.313</td>
</tr>
</tbody>
</table>

\[ \sum w_j \text{ for all } w_j > 0.01 = 1 \]

Observe that only \( w_j > 0.01 \) is presented. Thus, if sum of all presented \( w_j \) is below one this implies that some regions are assigned with weights under 0.01 but above zero.

### Table A8 Weights Employment rate (Regions dataset)

<table>
<thead>
<tr>
<th>Region</th>
<th>( w_j )</th>
<th>Region</th>
<th>( w_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td></td>
<td>Lüneburg (Germany)</td>
<td>0.218</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>Estonia</td>
<td>0.162</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>Déľ-Alföld (Hungary)</td>
<td>0.011</td>
</tr>
<tr>
<td>Unterfranken (Germany)</td>
<td>0.026</td>
<td>Wien (Austria)</td>
<td>0.01</td>
</tr>
<tr>
<td>Flevoland (Netherlands)</td>
<td>0.188</td>
<td>Lincolnshire (UK)</td>
<td>0.011</td>
</tr>
<tr>
<td>Opolskie (Poland)</td>
<td>0.15</td>
<td>East Wales (UK)</td>
<td>0.014</td>
</tr>
</tbody>
</table>

\[ \sum w_j \text{ for all } w_j > 0.01 = 0.364 \]

Observe that only \( w_j > 0.01 \) is presented. Thus, if sum of all presented \( w_j \) is below one this implies that some regions are assigned with weights under 0.01 but above zero.

### Table A9 Weights Unemployment rate (Countries dataset)

<table>
<thead>
<tr>
<th>Country</th>
<th>( w_j )</th>
<th>Country</th>
<th>( w_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td></td>
<td>Lithuania</td>
<td>0.239</td>
</tr>
<tr>
<td>Hungary</td>
<td>0.724</td>
<td>Netherlands</td>
<td>0.096</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.062</td>
<td>Slovenia</td>
<td>0.177</td>
</tr>
<tr>
<td>Finland</td>
<td>0.214</td>
<td>Finland</td>
<td>0.487</td>
</tr>
</tbody>
</table>

\[ \sum w_j \text{ for all } w_j > 0.01 = 1 \]

Observe that only \( w_j > 0.01 \) is presented. Thus, if sum of all presented \( w_j \) is below one this implies that some regions are assigned with weights under 0.01 but above zero.
<table>
<thead>
<tr>
<th>Region</th>
<th>$w_j$</th>
<th>Region</th>
<th>$w_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td></td>
<td>Berlin (Germany)</td>
<td>0.012</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>Schleswig-Holstein (Germany)</td>
<td>0.012</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>Illes Balears (Spain)</td>
<td>0.175</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>Flevoland (Netherlands)</td>
<td>0.01</td>
</tr>
<tr>
<td>Arnsberg (Germany)</td>
<td>0.126</td>
<td>Surrey (UK)</td>
<td>0.01</td>
</tr>
<tr>
<td>Kentriki Makedonia (Greece)</td>
<td>0.04</td>
<td>Espace Mittelland (Switzerland)</td>
<td>0.01</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.057</td>
<td>Istanbul (Turkey)</td>
<td>0.01</td>
</tr>
<tr>
<td>Észak-Margyarország (Hungary)</td>
<td>0.216</td>
<td>Kayseri (Turkey)</td>
<td>0.011</td>
</tr>
<tr>
<td>Észak-Alföld (Hungary)</td>
<td>0.335</td>
<td>Zonguldak (Turkey)</td>
<td>0.01</td>
</tr>
<tr>
<td>Ostösterreich (Austria)</td>
<td>0.226</td>
<td>Malatya (Turkey)</td>
<td>0.065</td>
</tr>
</tbody>
</table>

\[ \sum w_j \text{ for all } w_j > 0.01 = 1 \]

Observe that only $w_j > 0.01$ is presented. Thus, if sum of all presented $w_j$ is below one this implies that some regions are assigned with weights under 0.01 but above zero.
**AT1 Further explanation and derivation of DiD-estimator**

To clarify the mechanism in its simplest form, consider two periods, one pre-treatment and one as treatment is presumed to be efficient, referred to as treatment period. Consider further two groups, one treatment group and one control group. Let $P$ be equal to zero for the first period and one for the second, i.e. the treatment period. Further, let $G$ be equal to zero for the control group and one for the treatment group. Then apply these definitions to the model specified above to the definition of the DiD-estimator underneath.

$$E(y | P = 0, G = 0) = \beta_0$$
$$E(y | P = 1, G = 0) = \beta_0 + \beta_1$$
$$E(y | P = 0, G = 1) = \beta_0 + \beta_2$$
$$E(y | P = 1, G = 1) = \beta_0 + \beta_1 + \beta_2 + \delta_{DD}$$

The change between the first and the second period for control group is defined by:

$$\beta_0 + \beta_1 = E(y | P = 1, G = 0) - E(y | P = 0, G = 1)$$

The change between the first and the second period for treatment group is defined by:

$$\beta_0 = E(y | P = 0, G = 1) - E(y | P = 0, G = 0)$$

The difference in difference is therefore defined as:

$$\delta_{DD} = E(y | P = 1, G = 1) - E(y | P = 0, G = 1) - E(y | P = 1, G = 0) - E(y | P = 0, G = 0)$$

However, since this is an estimation and not a parameter it should be changed to its sample analogous, i.e. the DiD estimator:

$$\delta_{DD} = (\bar{y}_{t=2}^{TG} - \bar{y}_{t=1}^{TG}) - (\bar{y}_{t=2}^{CG} - \bar{y}_{t=1}^{CG})$$