REVISITING THE UNION – NON-UNION WAGE DIFFERENTIAL IN HUNGARY

By

Ioannis Gutzianas

Submitted to
Central European University
Department of Political Science

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Tamás Rudas

Budapest, Hungary

2022

Abstract This thesis deals with the estimation of the collective bargaining wage premium in Hungary. As such the study concentrates on what is referred to as the monopoly face of trade unions but tries to relate this to a wider understanding of how unions work. Previous research has shown a small and declining wage premium; thus, a continuation of this trend could pose serious questions about the legitimacy of Hungarian unions. This possibility also has wider implications given the interconnectedness of the faces of unions, meaning a small wage premium could restrict the possibilities of these organizations in the political arena as well. Utilizing a causal inference framework, the effect is estimated through both linear regression and matching using an employee-employer linked dataset. The initially observed raw wage gap of 27.3% decreases to between 1.5% and 2.2% after the estimation procedure, which fits in with the previously observed declining trend. Furthermore, these estimates likely suffer from an upwards bias due to the presence of unobservables, which means that the real effect is likely even lower.

Keywords union wage premium, collective bargaining wage premium, labour market, trade unions, matching, CBPS

Acknowledgements

First, and foremost I would like to thank my thesis supervisor, Professor Tamás Rudas, who agreed to be part of this project at a time when it did not necessarily look like something that is actually going to happen. Furthermore, he provided me with valuable feedback that changed this thesis for the better.

I also want to thank everyone who helped me from ITM and Enikő and Maja from KSH for their tremendous help in getting hold of the necessary data that made this research possible.

I am grateful to all my classmates and professors who made my experiences at CEU much more challenging and fun, especially my good friend Toni, without whom this would have been a very dull affair.

Last, but not least I want to thank my friends and family for supporting me all my life and putting me in a position where I had the chance to attend such a renowned university. Special thanks go out to my girlfriend Nóri who made sure that I snap out of my laziness and finally write this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Protecting and furthering the interests of workers is, and always has been, at the core of the labour movement. The last twelve or so years have seen significant changes to the institutional environment that trade unions face in Hungary, where three consecutive right-wing governments, ruling with a supermajority, have implemented two major reforms to the country's Labour Code (Bíró Nagy, Kadlót and Varga, 2012; Kertész, 2020). The first reform was a complete redraft that weakened the position of unions and strengthened employers, while the second one made workers remarkable more vulnerable and was dubbed the "Slave Act". At the midst of this hostile environment, it is fair to wonder to what extent Hungarian trade unions can achieve success in their work. There are a number of ways the success of labour unions can be defined, but this thesis is specifically interested in their ability to influence the wage setting process.

Using different econometric strategies this thesis will attempt to estimate the wage premium that is achieved through the means of collective bargaining. Continental European tradition refers to this as the collective bargaining wage premium, which shows the wage differential between workers who are covered by a collective bargaining agreement and those who are not. The thesis will also attempt to address certain methodological questions with regards to estimating this wage premium, specifically what role unobservables have in determining both wages and union status.

Exploring the effect that trade unions have on wages has been one of the most extensively studied topics of labour studies. The origins of the research program can mainly be found in North America, and it quickly gained popularity in most Anglosphere countries (Ashenfelter, Layard and Card, 1986; Crockett and Hall, 1986; Lewis, 1986; Freeman and Medoff, 1984). This is not to say that other parts of the world did not adopt this topic in some form or another,

but the geographical distribution of such research clearly skews towards the aforementioned region (Choi and Ramos, 2021; Wittenberg and Kerr, 2019). Continental Europe has a deep history of researching the effects that trade unions have on society (Esping-Andersen, 1990; Iversen, Pontusson, and Soskice, 2000) and even though the institutional setting differs significantly from those of the US, UK or Australia, the collective bargaining wage premium has received significant attention from scholars too (and de la Rica, 2006; DiNardo and Lee, 2004).

The case of Hungary, and all other nations from the former Eastern Bloc, is an interesting one in this regard. After the fall of the Soviet puppet regimes the environment that organized labour operated in changed radically, wages went from being set by a central authority to being the subject of bargaining and unions went from being accessories of the system to actually having to fight for their members, and workers in general. Despite this, not much attention has been devoted to the collective bargaining wage premium in Hungary, only a handful of papers have been written on the subject (Neumann, 2001; Neumann, 2006; Kertesi and Köllő, 2003b; Rigó, 2008). The fact that no scholarly attention has been paid to this topic in the last decade is especially concerning in the lights of the reforms that were mentioned above.

Studying and institutionalizing the needs and grievances of people who on an individual basis do not hold significant, if any, political and economic power is important if we want a well-functioning society. There are a number of ways in which we can approach this problem, but a historic and proven institution that has fared quit well in this regard is organized labour, specifically trade unions. It is no secret that unions are not experiencing their golden years in almost any aspect. It is also impossible to tell whether this is a periodic slowdown or society has changed so much that unions are unable to fulfil the role they once did. For those who still believe that unions have a key role to play in the future, it is important to conduct studies such

as this one, so the discussions about the remedies to this situation can be based on empirical work rather than intuitions.

Outline

The thesis is organized as follows. First, I am going to discuss trade unions and the institutional environment they face in Hungary, as well as the rich research program that concern the estimation of the above-mentioned wage premia (Chapter 1). Then I will present the data and the relevant methodologies (Chapter 2). I will start out by describing the different data sources that I consulted, and the steps that were taken to achieve the final data set that was used for analysis which will be followed by an extensive discussion of the two main methodologies of this thesis, linear regression, and matching. Third, I will present the results of the statistical analysis and discuss some potential interpretations (Chapter 3). The conclusion will summarize the findings and will outline some potential avenues for future research (Conclusion). The appendices will contain supporting material that I was not able to include in the thesis due to its various limitations, including further information on data sources, additional numerical and graphical results of the statistical analysis and the R Studio script as well.

Chapter 1 – Trade unions, institutional environment, and the wage premium

Although as I already stated above, this thesis deals with the issue of wage premia arising from the presence of Collective Bargaining Agreements specifically in Hungary this is virtually impossible without discussing trade unions. Although legally, it is possible to make Collective Bargaining Agreements without the involvement of unions, out of the more than 5200 such documents that were uploaded into the government's central database, only eight such cases can be found, none of which are still in force. Thus, I will commence with discussing trade unions in general, and specifically in Hungary. Following this I will highlight the institutional arrangements that the country's unions operate in. Finally, I will show some of the more important aspects of the union wage premium research program and its continental offspring, the collective bargaining wage premium.

1.1 Trade Unions

Organized labour, and as such trade unions, have been with us since the early days of industrialization, as the first strike that was recorded happened in Philadelphia in 1786 (Selig, 2006). Both the world and trade unions have changed radically ever since then, but at their core they remain the main tool in the hands of workers to enforce their needs and demands. The International Labour Organization (ILO) of the United Nations (UN) defines trade unions as "a workers' organization constituted for the purpose of furthering and defending the interests of workers" (ILO, 2022), which leaves a lot of room for interpreting their concrete role both in industrial relations, but in society as well.

Freeman and Madoff (1984) distinguish two "faces" of unions: the so-called monopoly face which refers to the union's power in the wage setting process, and voice/response face, which represents the other workplace obligations of unions which concern protecting workers'

interests. The latter involves things like ensuring a safe work environment, reasonable working hours and days or protecting workers from layoffs.

The former is the real interest of this thesis and is also referred to as the rent-seeking ability of unions (Conolly, Hirsch and Hirschey, 1986; Bronars and Lott, 1989). Rent-seeking, which refers to "rewards and prizes not earned or not consistent with competitive market returns" (Congleton, Hillman, and Konrad, 2008 pp. 1), in this case translates to: unions receiving parts of the profit without adding to productivity. That is only partially true in our case because most of the extracted rents go to workers who are the primary drivers behind profits, but for the sake of consistency with the existing literature I will keep referring to this as rent-seeking.

Almost all economists since then have based their conceptualization of issues regarding trade unions based off this distinction (Hirsch, 2004). I find this to be a grave error and a good example of how there is a strong need for "cross fertilization" amongst social sciences. What I allude at is how the above dual character of trade unions neglects an especially important aspect of these institutions: their political face.

Under political face, I mean the work unions do to influence parts of the political process, from simple policy decisions to elections. This face was more pronounced historically, but the fact that it got weakened is no coincidence. One of the important reasons why unions were able to become the leaders of the labour movement and as such, took a leading role in class struggle, is that they were able to have a strong presence in the most important political institutions (Upchurch, Taylor, and Mathers, 2009). Many of the major labour/social democratic parties of the western world were either branches of or existed in a sort of symbiosis with major trade unions of their countries, such as the German Social Democratic Party or the British Labour Party (Kitschelt, 1994). Obviously, the weakening of class consciousness and through that, the disappearance of class struggle as a political project cannot solely be the responsibility of the

labour movement, with all the changes that have happened in society their job was not an easy one (Éber, 2020). Nevertheless, to understand developments of the union movement we have to look at all three faces.

Why have I devoted considerable time and space to discuss the political face of trade unions, when clearly, this thesis concerns the first, rent-seeking, face of these organizations? Because these faces do not exist in separation from each other, but rather, they form a symbiotic relationship and the labour movement's performance in any of the three faces has deterministic consequences for its performance in the others. A union that is unable to negotiate for higher wages or ensure a healthy work environment will never have the membership to exert power on a higher level. At the same time, unions that fail to influence political actors will face a hostile legal and institutional environment.

1.2 Trade Unions in Hungary

Printers founded the first trade union in Hungary in 1862, 14 years after they secured the first Collective Bargaining Agreement in the country's history (Lux, 2008). The following 80 years of their history was spent working in a sort of symbiosis with the Hungarian Social Democratic Party and culminated in some major achievements. At their height, these institutions had memberships in the hundreds of thousands and after the Aster Revolution of 1918 they took over many functions of the state, primarily in planning and supply chains (Hatos, 2018). Following the fall of the Hungarian Soviet Republic in 1919 and the resulting return of the Kingdom of Hungary, organized labour never recovered to their pre-revolutionary strength and after World War II and the Soviet takeover of the country, they got incorporated into the Hungarian Working People's Party.

Between 1948, the aforementioned takeover, and 1989, the fall of the soviet style system, trade unions did exist in name, but not in function (Lux, 2008). Most of their functions were

representative or of little consequence, like arranging holidays for workers. After the democratisation of the country followed the establishment of the market economy and multiple phases of privatization, where workers' interests were supposed to be protected by unions who have not engaged in such work for more than forty years. Although some scholars have noted that, especially between 1994 and 1998 when the Hungarian Socialist Party was in government, trade unions were successful in influencing the budgeting decisions of the state (Tóth, 2000), there were only sporadic successes in defending workers from the consequences of privatization (Héthy, 1994; Neumann, 1997).

In the last twenty years there were only six when more than 10,000 workers took part in strikes and only four when more than ten strikes were called (KSH, 2021). Figure 1.1 also highlights some of the problems of trade unions in Hungary.

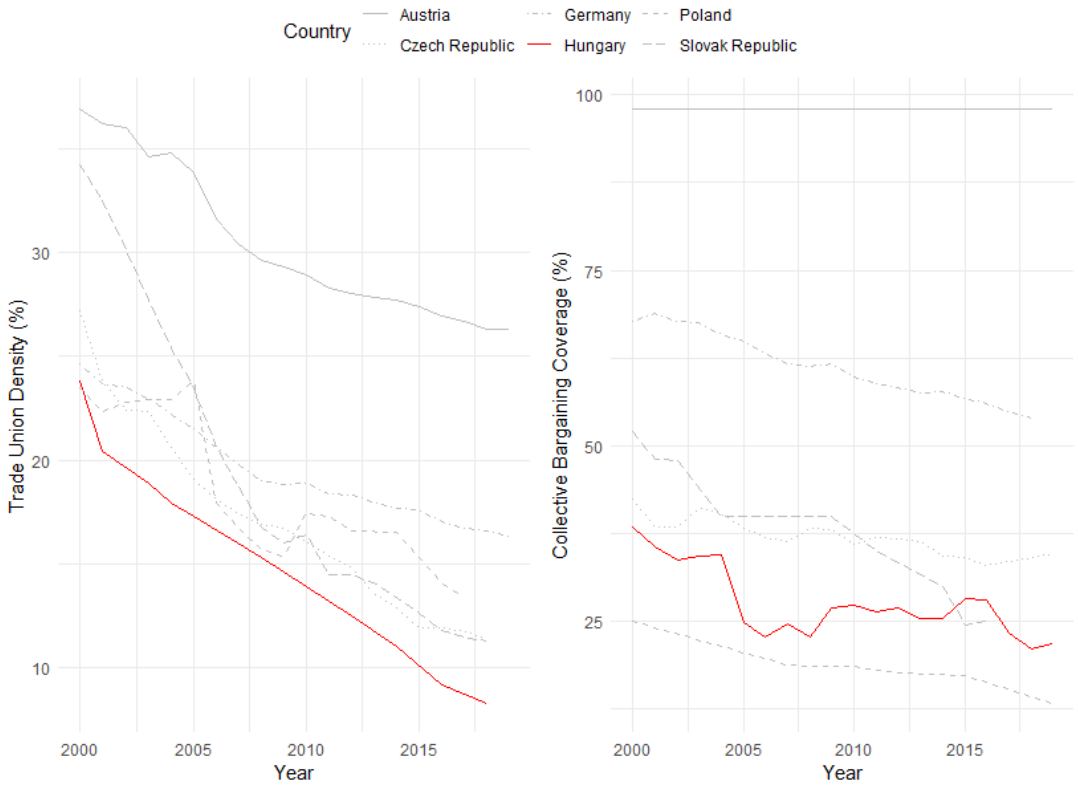


Figure 1.1- Trade Union Density and Collective Bargaining Coverage in Hungary and other countries 2000-2019 Source: OECD

Note: Missing values were interpolated/extrapolated using R

Both union density and collective bargaining coverage had a negative trend in the last twenty years in Hungary and in most other countries as well. While the decline in the latter is certainly concerning, after all, it is above 15 percentage points, but it does not stand out from the wider trends, at the same time, the decline in the former is quite breath-taking. While union density was around the regional average and comfortably in the double digits, slightly below 25%, in 2000 in merely 10 years it declined to well below 10%. Union power and the performance of unions cannot be evaluated simply on these measures, but they show continuous weakening and that the country ranks amongst the bottom performers, not just regionally, but in fact, in the whole European Union.

Ever since the country's democratisation and even more so nowadays, the Hungarian trade union movement is very decentralized and fragmented. Currently there are four (plus one) national trade union confederations that take part in the tripartite forum for consultation between government, employers and trade unions: the Trade Union Association of Intellectuals (*Értelmiségi Szakszervezeti Tömörülés*, ÉSZT), the Democratic League of Independent Trade Unions (*Független Szakszervezetek Demokratikus Ligája*, LIGA), the National Alliance of Workers Councils (*Munkástanácsok Országos Szövetsége*, MOSZ) and the Hungarian Trade Union Association (*Magyar Szakszervezeti Szövetség*, MSZSZ), which is closely associated with the Co-operative Forum of Trade Unions (*Szakszervezetek Együttműködési Fóruma*, SZEF). I should also note that there are a number of independent trade unions as well, for example the Independent Trade Union of Audi Hungária which is considered to be the most successful Hungarian trade union. Although there is some degree of cooperation between these unions, political and personal conflicts prevent effective work on the national level (ETUI, 2022).

In terms of political connections, there are two confederations that stand out. On the one hand, I already discussed how unions were part of the system before the democratization, and not all

those institutional arrangements disappeared after 1989. MSZSZ and its predecessors had historically cordial relations with the successor party of the ruling party of the pre-eighty-nine system, the already mentioned Hungarian Socialist Party. When the aforementioned party won the elections in 1994 some trade unionists got into parliament from their list, but even this did not prevent some of the biggest and most regressive austerity measures in the country's history (Scheiring, 2019). Ever since the downfall of the socialist party after 2010 and the weakening of the post-socialist left in Hungary, these relations have been less pronounced. On the other hand, MOSZ has traditionally defined itself as the Christian trade union and as such has had close relations with the political right in Hungary (MOSZ, 2022).

I have spent some time discussing how trade unions work and how they have evolved, or rather devolved, in Hungary in the last couple of decades, but now it is time to turn our attention to the environmental factors, that is the wage setting environment the legal system dictates.

1.3 Institutional background of the wage setting process

It is important to discuss some legal aspects of the wage setting process because these arrangements differ in a fundamental way in different countries, some of the most developed economies have completely different wage setting arrangements. On the one end of the spectrum there are countries with completely decentralized wage setting which happens at the firm level, like the US, Canada, or Japan, while in the other end we see countries with centralized bargaining systems where unions and employer federations run the process, like the Nordic countries or Austria (Calmfors and Driffill, 1988). Naturally, there are countries with systems that are somewhere between the two ends, a particular group that warrants highlighting is the countries where bargaining takes place at the industry level, like Germany or Belgium. Be aware: the wage setting process is more extensive than just collective bargaining, central wage setting (i.e., minimum wage) and individual bargaining are also part of it, but the latter two will not be discussed in this thesis.

Before 1989, there was virtually no bargaining process and wages were set by the central government in Hungary. After the democratisation, the first Labour Code (LC) was passed in 1992, and like most of the modern Hungarian legal system it was inspired by its German counterpart (Hajdú, 2019). As such, the original goal of the LC was to designate collective bargaining agreements as the main regulatory tool of the labour market (Kun, 2019). This did not work out as the legislators intended, and there are multiple explanations as to why. Some scholars claim that the main reason is the relative dispositive nature of the law, meaning that when it came to actual collective bargaining departure from the law was permitted, but only in favour of the employee (Gyulavári and Kártyás, 2015). The claim is that this rigidity prevented the self-regulation of the labour market.

Another reason for lack of success with the 1992 LC is that even though it was modelled on its German counterpart, a crucial part of German collective bargaining, industrial and sectoral wage setting never became a central tool in Hungary (Kohl, 2015). The option is there for such agreements and even some sectoral social dialogue committees were set up, but no significant industry level Collective Bargaining Agreement was ever struck (Berke, 2015). According to Kun (2019) the heavy reliance on industry level bargaining might also be a heritage of the pre-1989 era.

In 2012 a new LC was drafted and enacted by Parliament in the midst of a major financial crisis where unemployment was remarkably high (Grindt, 2013). It is also important to keep in mind that this LC was drafted by a right-wing government that was elected with a two thirds majority in 2010, which although is commonly described as being economically pragmatistic, in terms of its stances on unions and the labour movement it is best described as neoliberal (Sebők, 2019). Accordingly, the new LC was a large step in the direction of making the country's labour market more flexible. One of the changes came in the form of changing what I previously described as a relatively dispositive clause to being absolutely dispositive, which means that

Collective Bargaining Agreements can deviate from what is laid down in law, contrary to the old LC, even to the employee's detriment (Kun, 2019).

Supposedly, this change was enacted so that strengthen both employers and unions in the negotiating process by reducing the regulative function of the state (Grindt, 2013). It is hard to accept that reasoning if we take a look at the wider picture. The new LC radically changed the right to strike by making firm level strikes much harder to legally organize and making general strikes virtually impossible (Balczer, 2016). Another change the new law brought was the enhanced authority of workers' councils, representatives that are elected by workers, specifically, their right to conclude collective bargaining agreements if there is no union that commands at least 10% support from workers (Gyulavári and Kártyás, 2015). This in itself seems like a positive step for the workers, but it also has the potential of discouraging joining a union, but as I mentioned before, so far very few such collective agreements were made. A last point to make here is with regard to the former tripartite forum for consultation, the National Conciliation Council (*Országos Érdekegyeztető Tanács*, OÉT), an institution that that was extremely limited in terms of its legal mandate, but still was a venue where they could show their political face, was dissolved altogether and was replaced by the National Economic and Social Council (*Nemzeti Gazdasági és Társadalmi Tanács*) that was stripped even of those few legal mandates.

Overall, we see a coordinated attempt to weaken the labour movement and undermine the powers of trade unions in Hungary. There are aspects of the wage setting process that changes significantly throughout the last 30 years, but there are two important things that stayed constant and are important to keep in mind for the next part: wage setting happens at the firm level and every employee at the firm is covered by the Collective Bargaining Agreement (with the exception of management in most cases).

1.4 The Union Wage Premium and the Collective Bargaining Wage Premium

Ultimately, the questions and themes that were discussed so far are the building blocks of the research program where the main aim is to estimate the union wage premium. This program's history goes back a long way, as far back as the 1950s. Organized labour and the environments it operates in have changed significantly, so the way it is researched had to change too. One thing that remained constant is what researchers look for when conducting this research, which is the degree to which unions are able to maximize their rent-seeking, or rather, the degree of success they can achieve in what I referred to previously as the first face of trade unions. When it comes down to it, estimating the difference between the wages of unionized versus non-unionized workers the right way, will show us the effect unions have on the wage setting process.

In this last paragraph I have referred to this research program as being centred around the union wage premium, but previously I have also referred to this as the collective bargaining wage premium, which might seem like an inconsistency. The reason for this is something I have already discussed at some length; different countries have vastly different legal environments for their wage setting processes. A particular kind of legal environment is typical of most Anglo-Saxon countries, and contrary to what was said about Hungary, in these countries collective bargaining agreements only cover employees who are members of the union. Consequently, this research program was launched in North America, thus calling it union wage premium was the most suitable and logical option. From here on out I will be using the region appropriate name for the wage premium in question.

The earliest papers on the subject, all of which were written about the US, were summarized by Lewis (1986) and Freeman and Madoff (1984), and they found that the estimates of the union wage premium range between 10 and 30 percent. Even the lower bound seems like a sizeable premium, but the upper bound is particularly astonishing, considering that it would mean that

unions can ensure their members make 30% more than someone who is not a member. More recent surveys, particularly Blanchflower and Bryson (2007) and Blackburn (2008), have observed that the premium has gone through a steady decline since the 1990s and by the mid-2000s it ranged between 15 and 20 percent. This is still a sizable difference between wages, but the difference between this and the earlier surveys is very apparent. Most recently, papers that used more advanced techniques have found no or small, in the low one digits, union wage premium in the United States (Lee and Mas, 2012; Frandsen, 2012).

Canada seems to have gone down a similar path to its southern neighbour, Renaud (1998) in his survey of the earlier studies found that the union wage premium ranges between 10 and 25 percent. This seems to be a bit below what was found in the United States, but still, it is a notable wage difference. Canada also faced a steady decline following the 1990s, Fang and Verma (2002) noted that by the turn of the twenty first century the union wage premium declined to around 7%. In his 2018 paper Campolieti estimated that it has declined to under 5% by 2014. The picture is very similar in the United Kingdom where early estimates put the wage premium at around 20%, but more recent studies have found it to be low or negligible (Blanchflower and Bryson 2004).

In continental Europe, most studies that tried to estimate the collective bargaining wage premium are more recent and thus we can only discuss what was revealed in the last twenty or so years. Even amongst these studies there seems to be a great deal of variation not just between the different countries, but between the different estimates for the same countries as well. In the case of the Netherlands for example no positive effects of bargaining coverage were revealed (Hartog, Leuven, and Teulings, 2002), while in Germany some studies found a sizeable wage premium of around 10 percent (Stephan and Gerlach, 2005) while other estimates only range around 1 percent (Antonczyk, 2011). Card and De La Rica (2006) found the wage premium to be between 5 and 10 percent in Spain. Another interesting variation that was exploited in the

aforementioned study showed that on average firm level contracts produce higher wage premia than industry level contracts.

Ever since the democratization of the country this issue has received very little attention from scholars in Hungary currently there are two papers that deal with estimating the collective bargaining wage premium that I know of. They will be discussed in more detail in later parts of the thesis, but overall, we see that they estimated it to be between 2 and 3 percent (Rigó, 2008; Neumann, 2001).

A lot of these sharp declines in the estimated wage premium can be attributed to the change, or rather evolution, in the methodology used in the aforementioned papers, but this too will only be discussed in detail in a later chapter.

Chapter 2 – Data and methodology

2.1 Data

There are two main sources for the data that I draw on for the analysis conducted in this thesis. The first of which comes from the 2019 Individual wages and earnings data collection (Egyéni bérek és keresetek adatfelvétel) which is an annual survey conducted by the Hungarian Statistical Office (Központi Statisztikai Hivatal, KSH). KSH sends out the surveys in two waves, first to private sector organizations, then to budgetary authorities (institutions that are publicly financed) both of whom have a legal obligation to fill the survey.

The surveyed institutions must provide a set of variables about both themselves and their employees, the result of which is a matched employer-employee dataset, which gives me the chance to take into consideration some variation when it comes to employers. One of the questions in the survey concerns whether there is a Collective Bargaining Agreement at the given firm, but said question also happens to be the only one where answering is optional. For this reason, data regarding Collective Bargaining Agreements is unreliable and unfit for analysis.

This meant that I had to complement the microdata with an external data source on Collective Bargaining Agreements. This was provided to me by the Ministry for Innovation and Technology (Innovációs és Technológiai Minisztérium, ITM) which maintains a registry of all Collective Bargaining Agreements in the country since 1986. The complete list of active Collective Bargaining Agreements contains 3052 entries, where each entry constitutes a registered Collective Bargaining Agreement by a Hungarian company. The dataset also contains information on the date of the agreement, the company where the Collective Bargaining Agreement was signed and the type of the agreement (there are nine distinct types

of Collective Bargaining Agreements, but in this thesis, I am not going to exploit this categorization).

Even though in accordance with current legislation Collective Bargaining Agreements and their termination must be registered, there is no legal consequence to not adhering to said law, thus the complete list is also unfit for analysis. After consulting with representatives of trade unions I restricted the data to agreements that were made or renewed after the beginning of 2010, which resulted in 743 remaining entries on the list, which seems to be a sizeable drop from the 1267 Collective Bargaining Agreements in 1998 as reported by Neumann (2001).

After combining the original data on Collective Bargaining Agreements with the data received from ITM, collective bargaining coverage was a bit below 13% for all workers and slightly above 14% for private sector employees. Both of those figures stand significantly lower than the official coverage in Hungary in 2019 as reported by the OECD, which stands at 21,8% (OECD, 2022).

The 2019 Individual wages and earnings data collection survey dataset contains demographic, educational, occupational and wage data, as well as data on their employers, on 925,062 individuals. In accordance with the relevant literature, I excluded workers, who are employed at companies that have less than 10 employees; workers in the agriculture, forestry, fishing, mining, and oil industries; workers, who worked less than 60 hours in the month of the survey; workers employed in public works; self-employed individuals, student workers and everyone else, who is not employed through an employment contract (Breda, 2014; Campolieti, 2018; Gürtzgen, 2016). After examining the dataset, I also had to make some additional filtering, due to faulty data entry, thus I had to exclude people who have worked zero hours a week at the time of the survey as well as people whose gross hourly salary was less than the gross hourly

minimum wage in 2019. After filtering the dataset had 897,708 entries, of whom roughly 240 thousand workers were employed in the private sector.

Due to the complicated nature of getting access to microdata from KSH, most research, this one included, must be conducted in the statistical office's own "research room," which is designed to make it impossible to export information without KSH's approval. This unfortunately, means that there is only a rather limited timeframe for conducting research and the researcher is ever so exposed to the capabilities of the computer. The latter proved to be especially challenging in my research, as some of the methods that I use require the software to do high level calculations, which proved basically impossible with a sample of over 240 thousand in the given circumstances. For this reason, I had to restrict my sample size and to solve this problem I drew a random sample of fifty thousand from the private sector entries. For the sake of simplicity, I summarised this process in Figure 2.1.

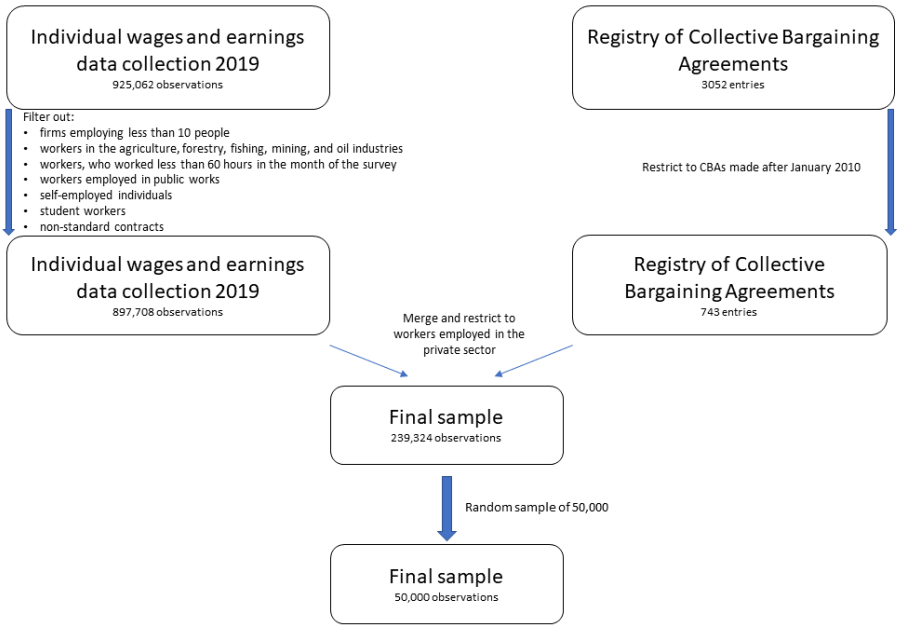


Figure 2.1- Data sources and the creation of the final sample

As it was expected, as trade unions are basically non-existent at firms that employ less than 10 people, collective bargaining coverage increased after filtering the dataset, to around 16%.

Although, I will discuss all the variables that I use in the analysis in more depth in the methodology chapter, I have already indicated that I use gross hourly wage as my variable of interest when it comes to wages. The Individual wages and earnings survey only collects information on monthly salaries, so to get the hourly wages I calculated them, by dividing each individual's gross monthly salaries with four times their usual paid work hours per week. This obviously does not give us the precise hourly wage of the workers, given that most months are not exactly four weeks, but since I will be estimating the percentage difference between the salaries of the two groups, this does not pose a significant problem in terms of the outcomes of my analysis.

Looking at the two main previous works on Hungarian collective bargaining wage premium, Neumann (2001) and Rigó (2008), it seems like both the availability and quality of data with regards to wages, unions and Collective Bargaining Agreements has significantly decreased in the last ten to fifteen years. From the mentioned works it seems like there was much stronger support for such inquiries by ministries (especially the since dissolved Ministry of Social Affairs and Labour) and the quality and quantity of data on Collective Bargaining Agreements was much better, given that said papers discuss details of these agreements, which are unavailable today.

2.2 Methodology

This thesis seeks to exploit some statistical methods to estimate the collective bargaining wage premium. Previous papers have used a wide range of such methods to identify said effect, but here I am going to concentrate on two of those: linear regression using Ordinary Least Squares (OLS) and matching. I will start by briefly, given its lesser importance, outlining my intentions and design with regards to the former, which will be followed by a more in-depth discussion of the latter, including its general theory, as well as the specific methods I used. Throughout I will

aim to highlight potential shortcomings of my methods of choosing as well as the issues caused by the lack of certain data sources that I have used.

2.2.1 Linear regression

2.2.1.1 Early approaches

Early studies that attempted the estimation of the union wage premium, used the strategy seen in equation 1 to do so:

$$(1) \qquad \log(wage)_{it} = x'_{it}\beta + \delta Union + u_{it}$$

where *i* stands for the individual, *t* for time and the dependent variable was an individual's *wage*, used in log form so the coefficient of interest can be interpreted as a semi-elasticity. Following that, *x'* is a vector of covariates that control for a number of demographic, socioeconomic and job characteristics. *Union* is a dummy variable, which denotes whether given individual is a member of a union and *u* is the error term. Such equations were used in the first studies that estimated the effect of unions on wages, and as I already discussed, yielded quite large estimates (Lewis, 1986; Farber, 1986; Freeman and Medoff, 1984).

Usually, the method of choice to estimate the above equation was linear regression using OLS, which I will simply refer to as linear regression in this case. In very simple terms, linear regression can be used to uncover the relationship between two variables, which cannot be explained by a third one (or a group of third variables). In my case this means that I want to isolate the part of the variance of wages that is due to the variance in union status, and for that I need to "remove" the variance that can be attributed to the other covariates.

The effect of any variables (or the polynomial form or interaction of included ones) that I did not include in the above equation will be included in the error term. In certain situations, this can cause significant problems for the researcher. One of the standard assumptions of linear regression is the exogeneity assumption, which requires that the variables in the equation are

unrelated to the error term. If this assumption is violated then we are facing omitted variables bias, which means that the coefficients we get are not the true ones.

The problem of endogeneity, which is another word for the violation of the exogeneity assumption, is deeply rooted in the union wage premium literature and it will come back again in this chapter. For now, I will concentrate on the critiques that were aimed at studies that utilized some form of equation 1. Most of the early criticism concerned the range of the included covariates (Lewis, 1986; Farber, 1986). Specifically, it was argued that simply by controlling for individual level variables, these models fail to account for a large part of what causes wages to vary, namely the firms they work for. We can confidently say that the size, financial situation and even the specific business area that a firm operates in has a significant effect on the wages said firm offers its employees.

It is important to point out, that firm level information was not necessarily left out of the mentioned papers because neglect from the side of the researchers, but rather that the availability of employee-employer linked datasets was not as widespread as today. Thus, the next generation of union wage premium research took a step forward and incorporated firm level data to their equations, after which the standard for them evolved to look as follows:

Expanding the equations with firm level data we get equation 2:

$$(2) \qquad \log(wage)_{ijt} = x'_{it}\beta + z'_{jt}\gamma + \delta Union_{it} + u_{ijt}$$

where everything stands for the same thing, and in addition we have the subscript *j* which denotes the firm, while *z'* is a matrix of firm level covariates. It can be argued that such an update vastly improves the reliability of our coefficients. Consequently, more or less this is the form of the equation used by Neumann (2001) in his already discussed paper on collective agreement premia. For the sake of comparability, the equation I use is something of an adaptation of the above one, as seen in equation 3:

(3) $$\log(Gross\ hourly\ salary)_{ijt} = x'_{it}\beta + z'_{jt}\gamma + \delta CBA_{jt} + u_{ijt}$$

As I already mentioned my dependent variable is the log of the *gross hourly wage* and *CBA* is the Collective Bargaining Agreement dummy. For individual level variables I have included: the person's gender (binary option in our case), their educational attainment classified into one of ten categories, the type of the employment contract, the form of employment and the period of service given in months. All the specific categorizations for these variables can be found in Appendix A. Firm level variables on the other hand are the number of employees at the firm as of September 2019 and the branch of the economy the firm operates in, as per the official KSH categorization.

One could still feel like something obvious is missing from this "list" looking at the above variables, and rightfully so. The level of wage a firm is able to offer largely depends on the financial situation said firm is in. Our dummy on the branch of the economy and the number of employees together can give us an indication on the potential revenue that a firm can generate, given that some sectors have access to more capital than others, but two seemingly similar companies in the same sector can look very different in terms of profitability. This problem could have been solved with a number of indicators, including cost and revenue figures or maybe the firm's EBITDA (Earnings Before Interest, Taxes, Depreciation) and alas a number of papers on the subject have done just that (Neumann, 2001; Wittenberg and Kerr, 2019). The answer to why this thesis fails to include any such variables is a simple one, including them in the dataset would have required a relatively large additional sum to be paid out to KSH, which at the point of my data request I could not finance.

2.2.1.2 Endogeneity

Earlier I have referred to linear regression as being of lesser importance in this thesis, which is only partially true. It is important in a sense because it gives us a clear point of comparison to

the effects identified by Neumann in 2001. The reason I consider matching to be the primary method is because the above strategy still suffers from a number of problems, most of which can be traced back to our earlier problem: endogeneity.

A concept that comes up quit frequently in research that deals with outcomes of human actions is that of the unobservable, which is arguably what still causes our estimation strategy to suffer from exogeneity. Even though the introduction of firm level variables greatly reduces the bias in the coefficients many scholars would still claim that due to unobservables likely having an effect on selection into treatment, bias is still present. After mentioning them so many times I should clear up what I mean under unobservables.

Under unobservables I mean the parts of a person's background that cannot be measured with conventional means that would make it suitable for a survey. A person's intrinsic, unobserved ability influences his or her life in many ways, from achievements and outcomes in school, to the quality of the job the individual can find and even retaining said job. Thus, based on what we already discussed, the effect of unobserved ability goes into the error term, which in turn becomes correlated with our treatment. This obviously means that the exogeneity assumption is violated and the coefficients are biased.

That being said, it is not evident what the direction of the aforementioned bias is and in the union wage premium literature has two conflicting answers to this dilemma. Generally speaking, positive correlation between the unobservable characteristics and our treatment, more specifically the union dummy, results in upwards bias, while negative correlation between the same variables gives us downwards bias (Rigó, 2008; Eren, 2007). Lewis (1986) argues that since the existence of a union at a given firm increases costs, that firm will face extra incentives to hire workers with higher skills and of higher quality, in order to upset the increase in costs with an increase in productivity. Another explanation for the possible wage premium, put

forward by Duncan and Stafford (1980), is that said premium is compensation for disadvantageous working conditions.

On the other hand, there is an argument for negative selection. Robinson (1989) approached the problem from the side of incentives and argued that the workers with the lowest wage, and thus skills, should be the most likely to join a union. This would mean that our coefficient on the CBA dummy incorporates the wage impact of said low skills, which again results in endogeneity.

For me personal, none of the arguments about negative selection have been very convincing. My issue with Robinson's (1989) argument goes back to its underlying logic, which presupposes a level of rationality which seems excessive to me. If individual workers were such rational actors, it seems unlikely to me that we would have seen the same rate of decline in union density all over the world, that we have actually witnessed. It might be obvious from this that I view the arguments for a positive selection more convincing and accordingly I expect the coefficient on the CBA dummy to have an upwards bias.

It should also be noted that, especially in the US context, it has been argued that on average linear regression has been acknowledged to produce reliable estimates of the union wage premium (Hirsch, 2004; Eren, 2007). The on average part refers to both being reliable over a longer time period as well as being reliable on average, meaning over all industries, but possibly not in specific ones.

2.2.1.3 Taking care of endogeneity?

Now that we have exhaustively discussed the problem with exogeneity, we should discuss how this problem can be overcome. It is obvious that linear regression is not suitable for tackling such problems, but there are other strategies for identifying the effect of Collective Bargaining Agreements without having to face endogeneity. One approach is estimating an equation that

looks more or less like equation 3 with fixed effects (Rigó, 2008). This approach is capable of controlling for selection bias, but for that it needs panel data, a dataset where variables are collected on the same individuals through multiple years.

The Individual wages and earnings survey was not designed with such research in mind, but according to the experts at KSH the available datasets could be used to construct such longitudinal data. That being said, the same financial restrictions that prevented the inclusion of firm level financial data apply here as well. I would argue that the lack of a time dimension poses a larger problem than that of financial data, not only because certain methodologies would require it, but rather because any collective bargaining wage premium value in a given year can be an outlier. To get a reliable picture of how Collective Bargaining Agreements influence wages in a country one should estimate said effect along multiple years.

Another way to circumventing unobservable bias is the Instrumental Variable (IV) approach (Ashraf, 1992; Mac Flynn, 2020). Without dwelling too deep into what IV is or how it works, IV requires the researcher to find an instrument that is a determinant of the presence of a Collective Bargaining Agreement but is strictly not a determinant of wages (Huntington-Klein, 2021). Few researchers have attempted finding such an instrument, amongst them Booth and Bryan (2004) included questions on workers perceptions of the effectiveness of the union presence in their respective firms, which they used as the instrument. On that I tend to agree with Bryson's (2002) critique of this kind of approach wherein he claims that such subjective questions about union effectiveness can hardly be considered as reliable predictors of union status.

Last but not least I want to mention the endogenous switching approach as used by Nahm, Dobbie and MacMillan (2017), again going into a detailed explanation of this method would go beyond the scope of this thesis, it suffices to say that this method uses simultaneous

estimation, much like IV, a method which has been found to produce large and unreliable estimates of the union wage premium on multiple occasions (Hirsch, 2004; Eren 2007; Blanchflower and Bryson, 2007).

Overall, we can see that there are a number of issues with regards to linear regression, but certain issues prevent the usage of the potential remedies, be it data availability, or my personal reservations. That being said, unfortunately, the prefect statistical method does not exist, so it is up to the researcher to make case for his/her best available solution to the problem. For obvious reasons I have not mentioned the approach which I previously described as the "primary" method of this thesis, matching, but I will get back to it in the next part.

2.2.2 Matching

In the following part I will discuss the basic idea behind matching as well as its two general forms. Following that I will explore the specific effect I am going to estimate, the Average Treatment on the Treated (ATT) and how this method fares when it comes to endogeneity. Finally, I will detail the specific matching methods that I used.

2.2.2.1 The general idea

The methods that we discussed so far attempted the estimation of the relative wage gap between unionized and non-unionized workers, or in our case workers with collective bargaining power and those without it. There is another approach to this, one that I am attempting in this thesis, which is using estimators that form estimates of counterfactuals. A counterfactual refers to an alternate scenario where we observe the individual of interest who received treatment, in a setting where he or she did not receive the treatment or vice versa (Campolieti, 2018). This sound immensely complicated, but it will be much easier to understand if I relate it to our topic. Say we are looking at the wages of workers who are covered by a collective bargaining

agreement, the counterfactual of this would be looking at the wages of those same workers, but in an alternate scenario where they are not covered by a collective bargaining agreement.

Research would be a significantly easier phenomenon if we were able to observe such alternate universes, but alas that is not the case. Thus, we are in a situation where we need to find a feasible substitute, and this is exactly where matching comes in. As opposed to regression which removes variation that is related to variation in the covariates, matching attempts constructing comparison groups that are similar according to a set of matching variables, thus there is no variation in said covariates, other than the treatment (Huntington-Klein, 2021). Making all variation disappear certainly does not sound like something that is feasible in any real-life scenario, thus the real goal is to get as close as possible.

2.2.2.2 The treatment effects

What I referred to as observing alternate universes is called the potential outcomes framework in the literature. We called them those because we only observe one of them for an individual at a given time (Eren, 2007). The specific treatment effect that I am looking for in this thesis is the average treatment on the treated (ATT), which shows the average treatment effect among those who actually received treatment as compared to if they never received it. (Huntington-Klein, 2021). This can be expressed as:

$$(4) \qquad ATT = E[w_1 - w_0 | CBA = 1, X] = E[w_1 | CBA = 1, X] - E[w_0 | CBA = 1, X]$$

where w_1 represents the wages of a worker covered by a collective bargaining agreement, while w_0 refers to the wages of a worker who is not covered by one and X is a set of observed covariates that affect both the treatment and the outcome. These are the two potential outcomes and the "$|CBA=1$" part is the condition that the worker has to be under a collective bargaining agreement.

The term "$E[w_1|CBA = 1, X]$" in equation 4 is simply the average wage of a worker under a CBA and can be observed in our dataset, while the second term, "$E[w_0|CBA = 1, X]$" is the average wage that said worker would receive if they were placed in a job that is not covered by such an agreement, and it is what we call a counterfactual. As we said before, we have no way of observing the counterfactual, so we need to construct it. It also has to be said that the relative wage gap that we estimate using linear regression, is not directly comparable to this ATT because, as Humphreys (2009) established, the former is bounded by the latter on one side and by the average treatment effect on the controls (ATC) on the other.

2.2.2.3 ATT bias

There are certain biases that can arise when researchers use sample means to construct their counterfactual, expressed as:

(5) $$E[w_0|CBA = 1, X] - E[w_0|CBA = 0, X]$$

where the first term is the one that we already saw in equation 4 as the counterfactual and the second term is the "constructed counterfactual" which we can observe in the data, thus the difference between the two is our bias. Heckman, Hidehiko, Smith, and Petra (1998) split the bias into three parts B1, B2 and B3. B1 arises when there is not sufficient overlap in the data, meaning for a sizeable portion of treated observations there are no comparable non-treated observations vice versa. We observe B2 when the distributions of X are different in the two populations. Finally, B3 is happens when despite having common support, what we called sufficient overlap, and successfully conditioning on observables there are differences, meaning equation 5 produces a non-zero result.

B1 and B2 are referred to as "bias due to observables" and we will be able to observe the extent to which they distort our estimated effects by checking the region of common support in the matched samples (B1) as well as assessing the quality of the matching procedures (B2). B3 on

the other hand is due to correlation between our treatment and unobserved factors and is called "bias due to unobservables" (Eren, 2007). Once again, we arrive at the issue of endogeneity and to assess its severity in our specific case we will have to discuss some of the assumptions that we make when using matching.

2.2.2.4 Assumptions for matching

In terms of these assumptions there are two main ones that we need to discuss. First, the one we have basically already mentioned: the common support assumption. Common support assumes that there are appropriate observations for every treated individual in the untreated group. I assume that in terms of population characteristics the common support assumption holds, and as I said previously, I will assess how much common support there is, but in any case, I am not going to restrict the matched sample to the parts of it that have common support, because that would then the estimated effect would no longer correspond to the ATT (Greifer, 2022).

The second assumption, called the Conditional Independence Assumption (CIA), states that, conditional on a set of observed characteristics, the ones I called X previously, the treatment and the untreated outcome are independent of each other. Selection into treatment based on the treated outcome does not have to be ruled out according to Heckman and Navarro (2004), so we can say that this is a bit more lenient compared to the linear regression assumption I previously discussed. Because I am using a large number of variables in my matching methods, some of which probably are correlated with unobserved variables, it is very likely that even in case the CIA is violated, the arising biases are lessened (Campolieti, 2018). Can we thus state that biases that are likely affecting our linear regression estimate are a non-issue when it comes to matching? I do not feel comfortable with making that statement, but I think it is step forward in both this and numerous other areas as well.

Now that we discussed the important theoretical consideration it is time to get an understanding of how matching constructs the counterfactual.

2.2.2.5 The big questions of matching

According to Nick Huntington-Klein (2021) there are five choices, or as we will see three plus two, a researcher has to make when he or she uses matching. These five are as follows:

1. What will our matching criteria be?

2. Are we selecting matches or constructing a matched weighted sample?

3. What is the worst acceptable match?

3+1. If we are selecting matches, how many?

3+2. If we are constructing a matched weighted sample, how will weights decay with distance?

As answers to these questions, we will have to define some concepts, so we can tell the software what it needs to look for. With regards to the first question there are two major schools of thought. First of them is distance matching, where the basic idea is that two individuals are familiar if they are close in values when it comes to our matching variables. Thus, our goal is to minimize the distance, but it is not trivial what measure of distance we are using when we match on more than one variable.

Although there are a number of appropriate measures of this, including the most commonly used Euclidean distance, I am going to use the one that is the most popular with researchers who use matching, the Mahalanobis distance (Huntington-Klein, 2021; Cunningham, 2021). I should also note that regardless of the distance measure we choose above a certain number of matching variables our research is going to suffer from the "curse of dimensionality." This basically means that as the number of matching variables increases, finding appropriate matches

becomes harder and harder (Mahony, 2014). Thanks to the large number of observations that the dataset has, this is not going to be a huge problem.

The other school of thought is propensity score matching. As opposed to distance matching, where we evaluate the distances for each individual matching variable, here similarity comes from having a similar probability of receiving treatment (Caliendo and Kopeinig, 2008). One can estimate said probability by running a logit or probit regression, there are other alternatives, but these are the most common, of their treatment on their matching variables and with the resulting coefficients they can calculate the probability of being treated. From there on out they just need to find the untreated observations which have the most similar probability to their treated observations, and they are more or less done. Although this seems like the far greater alternative because it takes care of the "curse of dimensionality," but, as always, there are some drawbacks here as well.

As I said earlier a researcher can estimate the probability of treatment with some sort of equation, generally probit or logit, but that is exactly where problems can arise. Traditionally the way to finding the right one is through iteration (Huntington-Klein, 2021). That means that researchers, through trial and error, must find the equation that works best for their project. King and Nielsen (2019) demonstrated through a simulation exercise that propensity score matching can decrease balance in the data, they call this phenomenon the "PSM paradox", which is largely due to this iterative nature of the method.

For our second question it is clear that we have two choices: selecting matches or constructing a matched weighted sample. The former one means we go about choosing our matches in a very stark way, where based on certain criteria, we decide if untreated observations get matched to any of our treated observations and those that do not, get discarded. The latter one is a lot like what I previously discussed, but instead of having a sharp in or out criterion as untreated

observations get farther from our treated observation, they get assigned smaller and smaller weights, until the weights become zero (Huntington-Klein, 2021).

The third question is a relatively simple one. I have already mentioned multiple times that we are trying to find untreated observations that are in some way close to a treated observation. To avoid the best match still being relatively far one can implement a bandwidth or a caliper, which are usually defined in terms of standard deviations of our matching variables, or the standard deviation of the propensity score. If no untreated observation is inside our selected caliper/bandwidth than that treated observation gets dropped from out sample.

The next two choices I have labelled as being additional ones, because they are conditional on previous choices we have made and my not even come up in certain research projects. When it comes to the right number of matches, we have three possibilities. First, we can choose to match exactly one untreated observation to each treated observation, this is called one-to-one matching. The second option is known as the "*k*-nearest-neighbour matching," which entails picking the k best matches, where the researcher has to decide on the exact value of *k*. Last, but not least, we can pick all the acceptable matches (Huntington-Klein, 2021). The way to deciding the scope of acceptable matches is what we already outlined when discussing choice number 3.

Another aspect of the right number of matches we have to discuss is whether we do matching with or without replacement. If we do the latter, then if an untreated observation gets matched to a treated observation it will not be matched again during the process. The former naturally entails that an untreated observation that gets matched remains available to be matched to further treated observations. There is no exact science to choosing one of the three options or choosing the value of *k* or deciding about matching with or without replacement, like many aspects of matching, finding the right choice for our exact project is an iterative process, but it is also down to a trade-off between variance and bias (Huntington-Klein, 2021).

With an increasing number of matches, by definition, the quality of every additional match is worse than that of the previous one, thus bias has to increase with that. At the same time with more matches also mean less sampling variation and so lower standard errors, which will make the estimates of the effect more accurate.

Our last choice requires us to choose a function which determines how weights will decay with distance. There are two main ways of going about this: using kernel functions and using inverse probability weighting (IPW). In the former one has to choose a kernel function, the most popular being the Epanechnikov kernel, which then gets "fed" a distance, the distance between the treated and untreated observation, and it returns a weight. The biggest possible weight gets assigned to the observations that have a distance of zero and then they decline as it moves away from said value. IPW on the other hand was specifically designed to work with propensity scores (Hirano, Imbens, and Ridder, 2003). The basic logic of it is: "it weights each observation by the inverse of the probability that it had the treatment value it had" (Huntington-Klein, 2021). This once again is a relatively confusing definition so I will present an example for clarity's sake. For an observation that was actually treated and has a propensity score of 0.75, which is above the 0.5 mark from where he or she is more likely to be treated than not, they will get a weight of 1/0.75. If another observation has the same propensity score, but was not actually treated, they will get a weight of 1/(1-0.75).

After the most important aspects of matching have been discussed it is time to explain the exact matching methods and variables that I use in this thesis.

2.2.2.6 MDM, CBPS and beyond

If the previous part taught us anything is that there is no single best practice for matching and it is not possible to predetermine what option is going to serve us best. For this reason, I am using multiple matching methods.

First, I will discuss my methods in terms of matching criteria. One of the routes I am taking is the standard Mahalanobis distance matching (MDM). As I previously explained, this method pairs observations based on the differences of their covariate values, and the one(s) with the least distance get paired. I will also utilise the other matching "family", PSM, as well. Because of the harsh critique that was levelled at said method, primarily by King and Nielsen (2019), researchers have been working on a solution that both improves covariate balance and is not reliant on an iterative process in finding a propensity score equation that works for the given project. Some such methods that were viable for this thesis include: genetic matching (Diamond and Sekhon, 2013), entropy balancing (Hainmuller, 2012) or the covariate balancing propensity score (CBPS) (Imai and Ratkovic, 2014).

Ultimately, I have decided to use the CBPS, because it is most like the propensity scores that are widely used (Campolieti, 2018). The CBPS belongs to the generalized method-of-moments framework and the basic idea behind it is to create a single model that both determines the equation that estimates the treatment assignment probability and optimizes the covariate balance, through the balancing weights. Compared to other similar models, the CBPS stands out because it achieves this dual goal without consulting the outcome data, which makes it more aligned with the original idea behind matching (Imai and Ratkovic, 2014; Rubin, 2007). This consequently elevates its applicability to be used in causal inference settings, including matching.

In terms of number of matches, in both cases we will look at multiple options, but because of the lacking computational power that I have already mentioned there was a limit to this. In the end I ran both the MDM and the CBPS models with the desired number of matches being: one, two, three and four respectively and in all cases the procedures were done with replacement. The former model is an example of selecting matches, while the latter one combines the two approaches, since it is a form of selecting matches, but they get assigned non-equal weights.

CBPS uses IPW for calculating sad matches (Imai and Ratkovic, 2014). Because the large sample resulted in good quality matches, there was no need for the use of a caliper or bandwidth.

In terms of matching variables, I am going to use the exact same variables that were included in the linear regression equation, with the exception of age and educational attainment. The reason for this is that every single one of the utilised matching methods decreased the balance when it comes to those two variables, so I have decided to exclude them from the procedures, but it was used at the effect estimation stage.

I have repeatedly talked about creating balance in the data with matching but assessing that is not a trivial matter. Thankfully, there are already established practises for this, mainly balance tables. Balance tables used for assessing balance before and after matching and, in the case of binary treatments, they show the standardized mean difference between the treated and untreated group for each matching variable. I will utilise balance tables to assess balance and in each case after said assessment was done, I will estimate the effect.

I used R (R Core Team, 2020) for conducting my analysis. Although using linear regression is part of the software's core package, I needed the MatchIt library (Ho, Imai, King, and Stuart, 2011) to not only do matching, but create visualizations based on the results. The aforementioned library contains the CBPS package (Fong, Ratkovic and Imai, 2021) as well which made it possible for me to use this method. I also need to mention the Stargazer package (Hlavac, 2018) which made summarising the linear regression results tremendously easier.

Chapter 3 - Results

In this chapter I will present the results of the different estimation procedures. First, as a basis for comparison I will show the raw wage premium that can be found in the original data. Following that I will discuss the results of the linear regression. After this I will get on to the different aspects of matching, starting with the assessment of the balance that our matching methods achieved. Consequently, I will briefly discuss the drivers of making collective bargaining agreements, as seen in the propensity score estimation. Lastly, I will take a look at wage premium estimates that the different matching methods have produced.

3.1 Raw wage premium

Most, if not all papers on the union wage premium/collective bargaining premium report some form of the raw wage premium (Rigó, 2008; Neumann, 2001; Kertesi and Köllő, 2003a; Eren, 2007; Mac Flynn, 2020). The raw wage premium is the difference between the average wages of workers who are covered by a Collective Bargaining Agreement and those who are not, without controlling for any covariates. This helps us later on as it serves as a point of comparison. As noted before the results of the linear regression and the ATT that matching produces cannot be directly compared, but they can both be compared to the raw wage premium.

In our case the raw wage premium is 27.3%, which means that on average the wages of workers who are covered by a Collective Bargaining Agreement are 27.3% more than of those who are not covered by one. If this were the actual wage premium its size would rival the size of the early estimates of the premium that I mentioned in the literature review. As we will see this is not the case and now, we can get on to discussing the results of the linear regression.

3.2 Relative wage gap as observed in the results of the linear regression

Table 3.1 presents the results of the linear regression.

	Dependent variable:				
	log(Gross_hourly_salary)				
	OLS				
	(1)	(2)	(3)	(4)	(5)
Constant	7.550*** (7.545, 7.555)	6.945*** (6.926, 6.963)	7.000*** (6.982, 7.018)	6.931*** (6.905, 6.958)	6.611*** (6.583, 6.639)
Collective Bargaining Agreement	**0.261*** (0.248, 0.274)**	**0.251*** (0.240, 0.262)**	**0.206*** (0.196, 0.217)**	**0.155*** (0.145, 0.166)**	**0.015** (0.003, 0.026)**
Gender (Base Male)		-0.165*** (-0.173, -0.157)	-0.138*** (-0.146, -0.130)	-0.101*** (-0.109, -0.093)	-0.115*** (-0.123, -0.107)
Age		-0.0002 (-0.001, 0.0001)	-0.001*** (-0.001, -0.001)	0.0004** (0.00005, 0.001)	0.001*** (0.001, 0.001)
Educational attainment		0.017*** (0.017, 0.018)	0.017*** (0.016, 0.017)	0.016*** (0.016, 0.017)	0.016*** (0.016, 0.016)
Fixed-term employment (Base Permanent employment)			-0.119*** (-0.138, -0.100)	-0.049*** (-0.068, -0.031)	-0.076*** (-0.094, -0.058)
Part-time contract (Base Full-time contract)			-0.200*** (-0.212, -0.188)	-0.173*** (-0.184, -0.162)	-0.160*** (-0.171, -0.149)
Subsidized full-time (Base Full-time contract)			-0.826*** (-0.883, -0.768)	-0.790*** (-0.844, -0.735)	-0.816*** (-0.869, -0.763)
Subsidized part-time (Base Full-time contract)			-0.534*** (-0.785, -0.283)	-0.526*** (-0.763, -0.288)	-0.498*** (-0.728, -0.269)
Period of service			0.001*** (0.001, 0.001)	0.001*** (0.001, 0.001)	0.001*** (0.001, 0.001)
Econ branch: Manufacture of textiles, clothing, leather, and leather products				-0.172*** (-0.211, -0.134)	-0.167*** (-0.204, -0.129)
Econ branch: Wood processing, manufacture of paper products, printing				-0.032* (-0.066, 0.002)	-0.020 (-0.053, 0.013)
Econ branch: Manufacture of chemicals and chemical products				0.185*** (0.135, 0.236)	0.214*** (0.165, 0.263)
Econ branch: Manufacture of pharmaceuticals				0.180*** (0.130, 0.230)	0.159*** (0.111, 0.208)
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products				0.161*** (0.131, 0.192)	0.129*** (0.100, 0.159)
Econ branch: Manufacture of basic metals and fabricated metal products				0.064*** (0.036, 0.093)	0.093*** (0.066, 0.121)
Econ branch: Manufacture of computers, electronic and optical products				0.132*** (0.097, 0.167)	0.075*** (0.041, 0.109)
Econ branch: Manufacture of electrical equipment				0.086*** (0.050, 0.122)	0.028 (-0.007, 0.063)
Econ branch: Manufacture of machinery				0.174*** (0.140, 0.208)	0.134*** (0.101, 0.167)

Econ branch: Manufacture of motor vehicles				0.273*** (0.245, 0.301)	0.169*** (0.142, 0.197)
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment				0.002 (-0.031, 0.035)	0.026 (-0.006, 0.058)
Econ branch: Electricity, gas, steam, and air conditioning supply				0.254*** (0.214, 0.295)	0.293*** (0.253, 0.332)
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination				-0.012 (-0.044, 0.020)	0.023 (-0.008, 0.054)
Econ branch: Construction				-0.098*** (-0.122, -0.073)	-0.007 (-0.031, 0.017)
Econ branch: Trade, vehicle repair				-0.005 (-0.027, 0.017)	0.052*** (0.030, 0.074)
Econ branch: Logistics				-0.046*** (-0.070, -0.021)	-0.034*** (-0.057, -0.010)
Econ branch: Accommodation services, hospitality				-0.178*** (-0.208, -0.149)	-0.126*** (-0.154, -0.097)
Econ branch: Information, communication				0.526*** (0.480, 0.572)	0.478*** (0.441, 0.507)
Econ branch: Financial and insurance activities				0.393*** (0.363, 0.423)	0.346*** (0.317, 0.375)
Econ branch: Real estate				0.009 (-0.026, 0.044)	0.095*** (0.061, 0.129)
Econ branch: Professional, scientific, and technical activities				0.377*** (0.334, 0.412)	0.427** (0.361, 0.486)
Econ branch: Administrative and support service activities				0.006 (-0.020, 0.031)	-0.015 (-0.039, 0.010)
Econ branch: Administration, defence; compulsory social security				0.021 (-0.052, 0.095)	0.007 (-0.064, 0.078)
Econ branch: Education				-0.258*** (-0.283, -0.233)	-0.159*** (-0.184, -0.135)
Econ branch: Healthcare				-0.118*** (-0.151, -0.085)	-0.051*** (-0.084, -0.019)
Econ branch: Social work				-0.176*** (-0.202, -0.150)	-0.093*** (-0.119, -0.068)
Econ branch: Arts, entertainment, leisure				0.020 (-0.018, 0.059)	0.068*** (0.031, 0.105)
Econ branch: Other				-0.104*** (-0.143, -0.065)	-0.062*** (-0.100, -0.024)
Number of employees					0.005*** (0.005, 0.006)
Observations	50,000	50,000	50,000	50,000	50,000
R^2	0.031	0.303	0.350	0.420	0.457
Adjusted R^2	0.031	0.303	0.350	0.419	0.457
Residual Std. Error	0.542 (df = 49998)	0.459 (df = 49995)	0.444 (df = 49990)	0.419 (df = 49958)	0.405 (df = 49957)
F Statistic	1,578.168*** (df = 1; 49998)	5,431.274*** (df = 4; 49995)	2,989.728*** (df = 9; 49990)	881.755*** (df = 41; 49958)	1,001.484*** (df = 42; 49957)

Note: Base for Econ branches is Econ branch: Manufacture of food products, beverages, and tobacco

*p<0.1; **p<0.05; ***p<0.01

Table 3.1- Results of the linear regression

In this table we see the coefficients of the different variables, that of the Collective Bargaining Agreement being highlighted in bold, through the five stages of the estimation. Here the stages simply represent waves of adding new variables, so we get a picture of how different kinds of variables influence the coefficient of the Collective Bargaining Agreement. Next to and below the coefficients we can see their confidence intervals, I opted to use these instead of the usual p values, because of the unhealthy obsession some researchers have about them. Still their significance levels can be deduced by the stars next to them.

The first stage is basically something that was already discussed, as it looks at the effect that Collective Bargaining Agreements have on the dependent variable, which is more or less the raw wage premium. Accordingly, the coefficient on the treatment is between 24.8% and 27.4% which corresponds to the 27.3% that I identified as the raw wage premium. Clearly this does not tell much, and the R squared is also showing us that not much of the variation is being explained by the explanatory variable. Interestingly enough in Neumann's (2001) estimation the collective bargaining dummy alone produced a similar coefficient, around 30%.

In the second stage three additional variables, gender, age, and educational attainment, get added to the estimation process, hence this is a stage where we expand with demographic characteristics. Even though the explanatory power of the estimation increases significantly compared to the previous one, rising from .031 to .303, the coefficient on the Collective Bargaining Agreement does not change that much. This stage puts the collective bargaining premium between 24% and 26.2% which differs only slightly from the previous estimate.

The third stage expands with three employment characteristics, the type of the employment contract, the form of employment and the period of service. These additions produce a heftier shift in the coefficient of the Collective Bargaining Agreement whereas it decreases to between 19.6% and 21.7%.

The next two stages are the ones where I introduce the firm level variables. First, at the fourth stage, I have added the variable on the different branches of the economy, which once again decreased the coefficient on the treatment to between 14.5% and 16.6%. Finally, at the fifth stage I add the number of employees, which is the one variable that incorporates the effect of the performance of a given company the best. Up to this point the introduction of new variables resulted in, at most, five percentage point decreases in consecutive stages. This changed drastically at the fifth stage where the introduction of the number of employees resulted in the coefficient on Collective Bargaining Agreement to decrease to between 0.3% and 2.6%.

This is a very significant decrease and it goes to show that a lot of the variation in wages that thus far we attributed to the presence of collective bargaining, is actually due to the size/performance of a given company. If the true coefficient is closer to the bottom part of that interval, that would mean that there is virtually no collective bargaining premium.

Comparing this result to the previous research programs that estimated some sort of relative wage gap in Hungary, mainly Neumann (2001), Kertesi and Köllő (2003b), and Rigó (2008), all of which use data from the late 90s and early 2000s, we get a rather alarming result. Neumann estimated that workers who are covered by a collective agreement earn 5% more on average, which declines to 2.7% when looking at firms who employ more than 300 employees, Kertesi and Köllő get similar results too. The most recent estimation was done by Rigó and already here we can see a decline compared to the previous two, as the researcher finds a relative wage gap of 3% using linear regression and 2% using a Fixed Effects estimation strategy.

Using the midpoint of our confidence interval, 1.5%, we can see that it is substantially lower than the effects that were identified beforehand. Compared to Rigó's estimation the collective bargaining premium declined by 50%, and by 70% if it is compared to the previous two estimates. If we take into account, the likely upwards bias that was discussed earlier it is safe

to say that the linear regression suggests that the wage premium associated with the Collective Bargaining Agreement has diminished if not disappeared by 2019.

3.3 The achieved balance as a result of matching

There are several ways a researcher can go about assessing the balance that the utilized matching methods were able to achieve. Most of them involve looking at the standardized mean difference between the same variables before and after matching and through that, the percentage improvement (or deterioration) in terms of balance in the data. Beyond that, as promised previously, I will also check Common Support and whether the distributions of the treated and untreated group are similar. First, I will present the balance in the original dataset, which I will follow up with doing the same for the different matching procedures.

	Standardized mean difference	Variance ratio (if applicable)
Female	0.1587	-
Male	0.1587	-
Permanent employment	0.2695	-
Fixed-term employment	0.2689	-
Full-time contract	0.0472	-
Part-time contract	0.0472	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.3835	2.1298
Econ branch: Manufacture of food products, beverages, and tobacco	0.0259	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0.0141	-
Econ branch: Wood processing, manufacture of paper products, printing	0.0204	-
Econ branch: Manufacture of chemicals and chemical products	0.0827	-
Econ branch: Manufacture of pharmaceuticals	0.1510	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0.0028	-
Econ branch: Manufacture of basic metals and fabricated metal products	0.0133	-
Econ branch: Manufacture of computers, electronic and optical products	0.1472	-
Econ branch: Manufacture of electrical equipment	0.1190	-
Econ branch: Manufacture of machinery	0.0983	-
Econ branch: Manufacture of motor vehicles	0.1424	-
Econ branch: Other manufacturing; installation and	0.0192	-

repair of industrial machinery and equipment		
Econ branch: Electricity, gas, steam, and air conditioning supply	0.1783	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0.1434	-
Econ branch: Construction	0.2444	-
Econ branch: Trade, vehicle repair	0.3000	-
Econ branch: Logistics	0.4156	-
Econ branch: Accommodation services, hospitality	0.3087	-
Econ branch: Information, communication	0.1372	-
Econ branch: Financial and insurance activities	0.0013	-
Econ branch: Real estate	0.1331	-
Econ branch: Professional, scientific, and technical activities	0.1691	-
Econ branch: Administrative and support service activities	0.1844	-
Econ branch: Administration, defence; compulsory social security	0.0491	-
Econ branch: Education	0.4596	-
Econ branch: Healthcare	0.1686	-
Econ branch: Social work	0.5935	-
Econ branch: Arts, entertainment, leisure	0.0889	-
Econ branch: Other	0.0506	-
Number of employees	1.5440	0.8968

Table 3.2 – Balance tables of the original dataset

In itself this table does not necessarily convey too much information about our data, but there are a few things we can conclude. First, let us discuss the two indicators of interest: standardized mean difference and variance ratio. The former is simply the difference between the means of the treatment and control groups after standardization, while the latter is the ratio of the variance of a variable in one group to that in the other (Austin, 2009). Variance ratios can only be calculated for continuous variables, hence the relative scarcity of them in this table. Furthermore, they are suitable for assessing whether what I referred to as B2 bias of ATTs are present in this paper, given that a variance ratio of one indicates that the variances of the samples are similar. That being said, in order to keep the thesis concise and digestible I am mostly going to rely on visual cues, like showing common support with density plots of propensity scores, plotting the continuous variables to evaluate their distributions and using so called love plots. Nevertheless, all balance tables will be available in Appendix B.

Our goals will be to get standardized mean differences as close to zero as possible and variance ratios as close to one as possible. Looking at this table we see that is most cases those goals seem far away, but also there is a lot of deviation in terms of the indicators. There are a couple of variables that have a rather low standardized mean difference, like the variables on employment contracts, and certain dummies on economy branches, but most others have significant differences in this regard. The number of employees variables specifically stands out, as it has by far the highest standardized mean difference, while having a variance ration that is relatively close to one.

3.3.1 Balance from MDM matching

First, l will demonstrate the balance that was achieved through the MDM procedures, Figure 3.1 is going to aid with that.

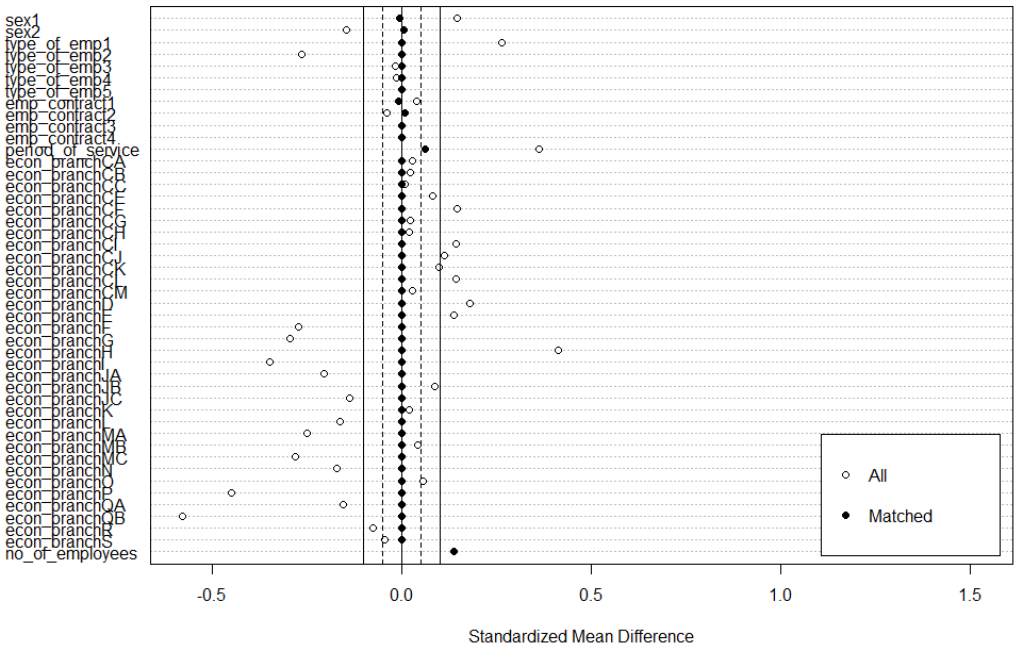


Figure 3.1– Change in standardized mean difference in the one-to-one MDM procedure

Figure 3.1 shows the love plot of the change in standardized mean differences as a result of the one-to-one MDM. Although the variables have their names from the software, they are the same

ones, in the same order as in the above tables. The dashed line next to the 0.0 point represent a standardized mean difference of 0.05, while the solid line next to it is 0.1. We can see that this method was able to achieve a standardized mean difference of at least below 0.1 for almost all the variables. with the exception of the number of employees variable, where said figure stands at 0.14. Even among the rest of the variables, only period of service remained over .05, at -062, the rest of them decreased to, or very close to, 0.

Because there are no stark differences between these MDM procedures the next graph is going to contain the love plots of the remaining three.

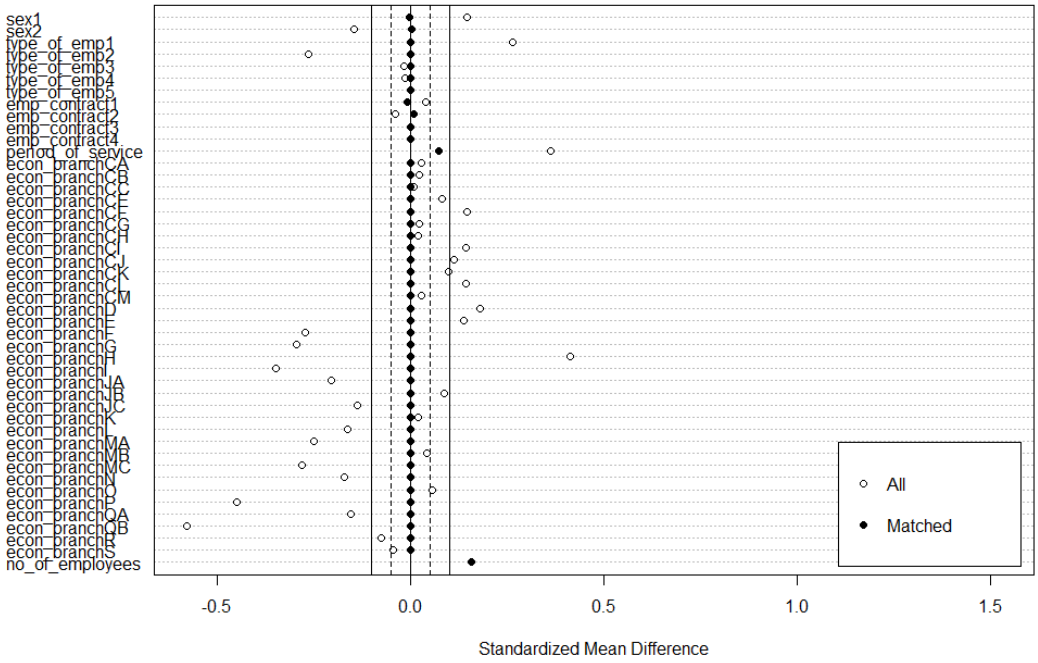


Figure 3.2 – Change in standardized mean difference in the two-to-one MDM procedure

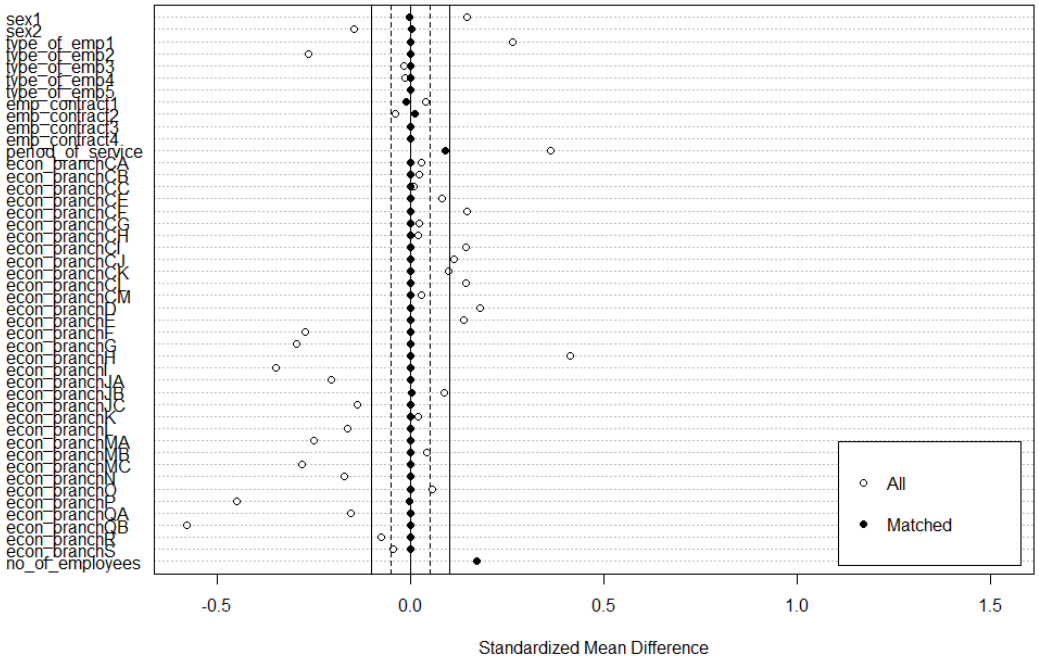


Figure 3.3 – Change in standardized mean difference in the three-to-one MDM procedure

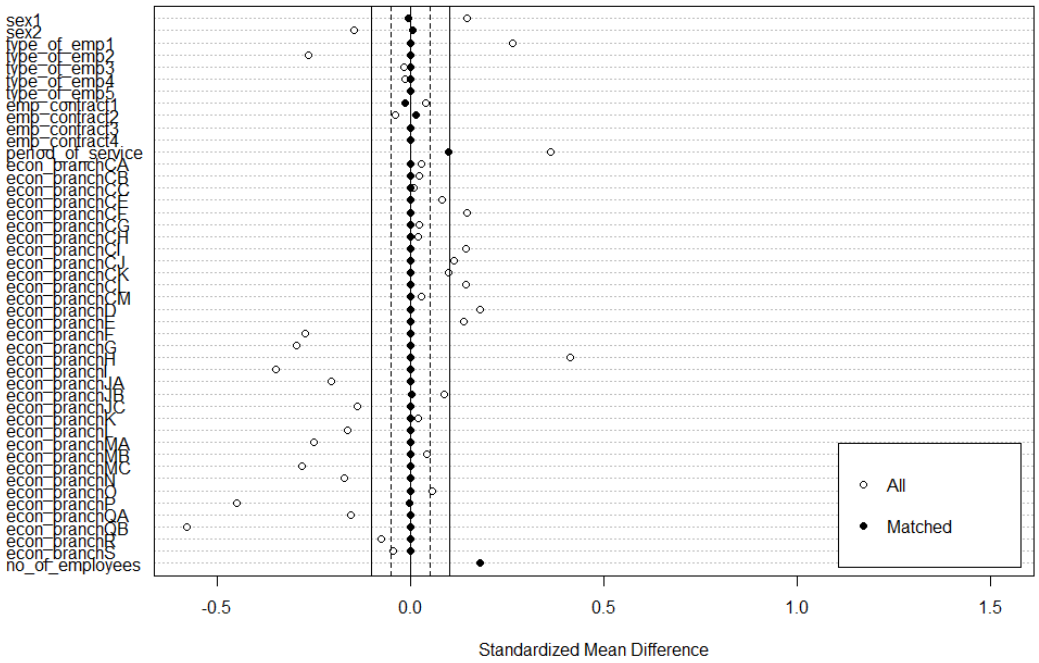


Figure 3.4 – Change in standardized mean difference in the four-to-one MDM procedure

What we see is a continuous increase in the standardized mean difference as the number of matches increases, specifically in the case of the continuous variables. In the case of period of service, it increased to .07 for two-to-one matching, .08 for three-to-one matching and .09 for four-to-one matching. We see similar deterioration with the number of employees variables too, increasing to .15 for two-to-one, .17 for three-to-one and .18 for four-to-one matching. This is not necessarily a huge problem, as I said earlier most of the choices one makes when he or she designs their matching criteria, this one included, is a trade-off between variance and bias, and all of our matching procedures achieve a satisfactory degree of balance improvement.

Table 3.3 is going to summarize the percentage improvement that the different matching procedures achieved with regards to the standardized mean differences.

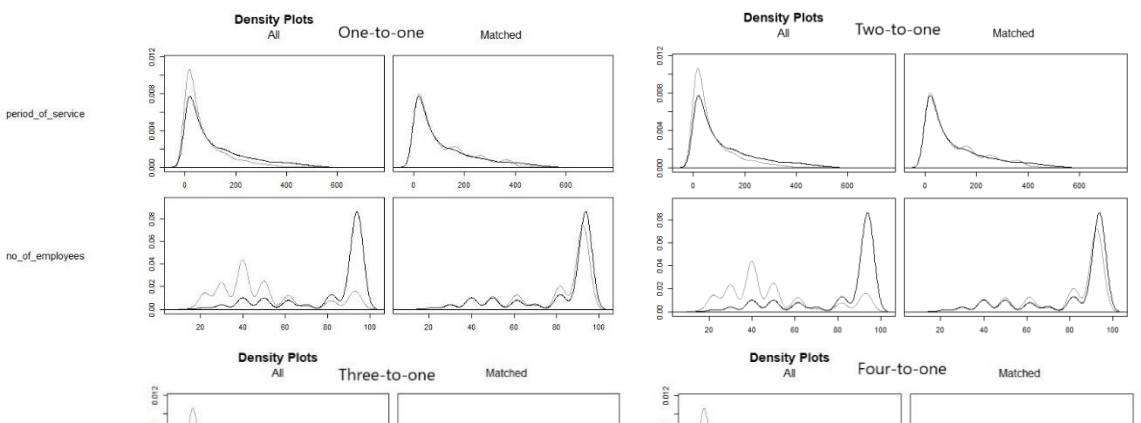
	One-to-one matching	Two-to-one matching	Three-to-one matching	Four-to-one matching
Female	95.7 %	95.7 %	98.1 %	96.7 %
Male	95.7 %	95.7 %	98.1 %	96.7 %
Permanent employment	100 %	100 %	100 %	100 %
Fixed-term employment	100 %	100 %	100 %	100 %
Full-time contract	80 %	80 %	71.1 %	66.2 %
Part-time contract	80 %	80 %	71.1 %	66.2 %
Period of service	82.8 %	79.5 %	75.6	73.0 %
Econ branch: Manufacture of food products, beverages, and tobacco	100 %	100 %	100 %	100 %
Econ branch: Manufacture of textiles, clothing, leather, and leather products	100 %	100 %	100 %	100 %
Econ branch: Wood processing, manufacture of paper products, printing	100 %	100 %	100 %	100 %
Econ branch: Manufacture of chemicals and chemical products	100 %	100 %	100 %	100 %
Econ branch: Manufacture of pharmaceuticals	100 %	100 %	100 %	100 %
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	100 %	100 %	100 %	96.9 %
Econ branch: Manufacture of basic metals and fabricated metal products	100 %	100 %	100 %	100 %
Econ branch: Manufacture of computers, electronic and optical products	100 %	100 %	100 %	100 %
Econ branch: Manufacture of electrical equipment	100 %	100 %	100 %	100 %
Econ branch: Manufacture of machinery	100 %	100 %	100 %	100 %
Econ branch: Manufacture of motor vehicles	100 %	100 %	100 %	100 %
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	100 %	100 %	100 %	100 %

Econ branch: Electricity, gas, steam, and air conditioning supply	100 %	100 %	100 %	100 %
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	100 %	100 %	99.9 %	100 %
Econ branch: Construction	100 %	100 %	100 %	99.9 %
Econ branch: Trade, vehicle repair	100 %	100 %	100 %	100 %
Econ branch: Logistics	99.9 %	100 %	100 %	99.9 %
Econ branch: Accommodation services, hospitality	100 %	100 %	100 %	100 %
Econ branch: Information, communication	100 %	100 %	99.1 %	98.3 %
Econ branch: Financial and insurance activities	100 %	100 %	100 %	100 %
Econ branch: Real estate	100 %	100 %	100 %	100 %
Econ branch: Professional, scientific, and technical activities	100 %	100 %	100 %	100 %
Econ branch: Administrative and support service activities	100 %	100 %	99.9 %	99.9 %
Econ branch: Administration, defence; compulsory social security	100 %	100 %	100 %	100 %
Econ branch: Education	100 %	99.8 %	99.5 %	99.2 %
Econ branch: Healthcare	100 %	100 %	100 %	100 %
Econ branch: Social work	100 %	100 %	100 %	99.8 %
Econ branch: Arts, entertainment, leisure	100 %	100 %	100 %	100 %
Econ branch: Other	100 %	100 %	100 %	100 %
Number of employees	91.1 %	89.7	88.8	88.2 %

Table 3.3 Percentage improvement in the standardized mean differences as a result of the different MDM procedures compared to original balance.

This table shows us a similar picture to the love plots. We can comfortably claim that the one-to-one matching offers the most improvement in terms of standardized mean difference and the only variables where we see significant differences are period of service and the type of employment contract. For the former we know that said difference was so close to zero to start with, that these deviations in the inspected improvement cause no problems in terms of balance creation.

To conclude the assessment of balance for the MDM procedures, let us look at the shape of the distributions of our continuous variables before and after matching, to make sure that we do not face a situation where B2 bias influences the results.



Like our previous tables and graphs, this one once again shows us that the MDM matching procedures were quite successful in achieving balance, but also, that balance was achieved with the treated and untreated observations having similar, if not identical distributions.

3.3.2 Balance from CBPS matching

I already made it clear that there is a philosophical difference between distance matching procedures and propensity score matching procedures, where for the latter is not based on minimizing the distance between the covariate values of treated and untreated observations. This suggests that looking at the same metrics, standardized mean differences, of the CBPS matching procedures does not make sense, which is a fair assessment. Despite this, I am going to briefly discuss how this method influenced balance, but only through recreating Table 3.4.

	One-to-one matching	Two-to-one matching	Three-to-one matching	Four-to-one matching
Female	59.2 %	54.9 %	77.9 %	70.4 %
Male	59.2 %	54.9 %	77.9 %	70.4 %
Permanent employment	92.0 %	94.4 %	94.7 %	94.0 %
Fixed-term employment	92.0 %	94.6 %	94.9 %	94.0 %
Full-time contract	33.3 %	0.8 %	20.0 %	10.8 %
Part-time contract	33.3 %	0.8 %	20.0 %	10.8 %
Period of service	64.8 %	72.8 %	71.7 %	75.9 %
Econ branch: Manufacture of food products, beverages, and tobacco	-42.3%	-26.2	-46.1 %	-38.9 %
Econ branch: Manufacture of textiles, clothing, leather, and leather products	-10.5 %	-32.1 %	-50.6 %	-56.2 %
Econ branch: Wood processing, manufacture of paper products, printing	-226.5 %	-251.7 %	-180.5 %	-217.1 %
Econ branch: Manufacture of chemicals and chemical products	70.7 %	72.0 %	64.8 %	68.8 %

Econ branch: Manufacture of pharmaceuticals	61.4 %	50.1 %	45.2 %	39.5 %
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	-59.6 %	-65.8 %	-30.0 %	1.0 %
Econ branch: Manufacture of basic metals and fabricated metal products	-145.7 %	-138.8 %	-120.3 %	-86.9 %
Econ branch: Manufacture of computers, electronic and optical products	69.7 %	79.0 %	74.1 %	73.0 %
Econ branch: Manufacture of electrical equipment	50.3 %	52.3 %	51.4 %	59.9 %
Econ branch: Manufacture of machinery	67.7 %	58.7 %	60.0 %	59.7 %
Econ branch: Manufacture of motor vehicles	79.8 %	65.1 %	68.5 %	66.2 %
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	97.0 %	69.8 %	43.5 %	33.5 %
Econ branch: Electricity, gas, steam, and air conditioning supply	-62.6 %	-71.3 %	-26.1 %	-41.6 %
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	81.4 %	79.0 %	31.9 %	51.0 %
Econ branch: Construction	99.7 %	95.8 %	96.9 %	94.8 %
Econ branch: Trade, vehicle repair	89.6 %	85.4 %	86.7 %	88.4 %
Econ branch: Logistics	86.7 %	84.3 %	85.3 %	84.1 %
Econ branch: Accommodation services, hospitality	86.7 %	89.6 %	91.2 %	90.5 %
Econ branch: Information, communication	87.7 %	88.1 %	95.3 %	87.5 %
Econ branch: Financial and insurance activities	78.4 %	49.7 %	90.1 %	-18.6 %
Econ branch: Real estate	83.4 %	93.8 %	42.5 %	93.3 %
Econ branch: Professional, scientific, and technical activities	88.4 %	86.6 %	83.6 %	84.9 %
Econ branch: Administrative and support service activities	86.5 %	83.5 %	87.4 %	88.1 %
Econ branch: Administration, defence; compulsory social security	49.3 %	53.5 %	51.2 %	47.9 %
Econ branch: Education	89.4 %	88.0 %	89.3 %	88.2 %
Econ branch: Healthcare	89.1 %	96.4 %	97.3 %	98.2 %
Econ branch: Social work	94.3 %	94.1 %	93.5 %	92.8 %
Econ branch: Arts, entertainment, leisure	80.4 %	80.4 %	79.2 %	80.9 %
Econ branch: Other	19.7 %	36.4 %	44.2 %	52.3 %
Number of employees	93.4 %	93.2 %	93.3 %	93.3 %

Table 3.4 - Percentage improvement in the standardized mean differences as a result of the different CBPS procedures compared to original balance

Comparing Table 3.4 to Table 3.3 we can see the philosophical difference I mentioned, as the former is much more hectic. We can see improvements that are comparable to what we saw in the latter table, but there are cases where there is virtually no improvement, where there is a slight deterioration, and even some, where the rate of deterioration is more than 100% or even surpassing 200%. Again, this is not necessarily a problem, but it goes to show that there is significant difference between these approaches.

Now for evaluating the CBPS matching, we will have to take a look at whether the Common Support assumption stands, that is whether there are available matches for all treated observations. To do that we will have to inspect the distribution of the propensity score for the treated and control groups. What we would like to see here is the two groups having similar distributions after matching is done.

Because all the distributions gained a similar shape after matching, I will only show the four-to-one CBPS matching in Figure 3.6 below, but each one will be included in Appendix C.

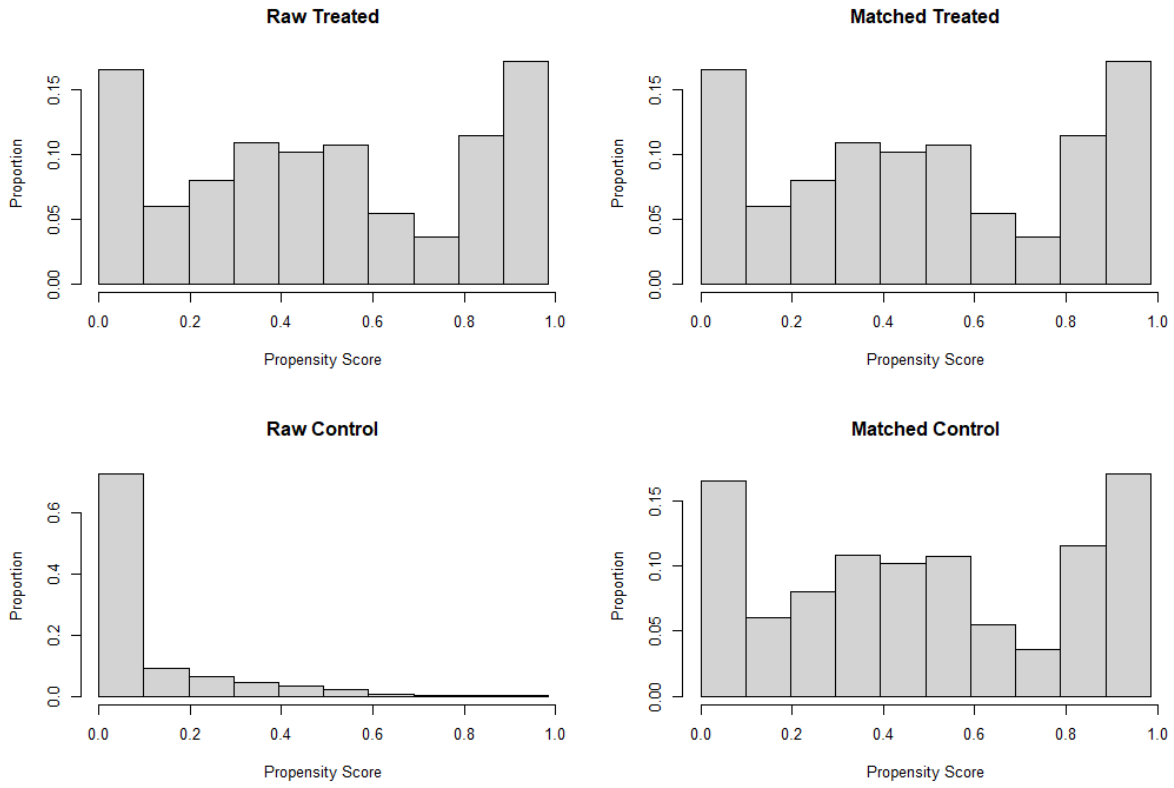


Figure 3.6 – The distribution of propensity scores before and after matching for the four-to-one CBPS procedure

Here the positive effect of having an unusually large sample size for matching shines through immensely. Before matching it seems like there are no suitable matches for the treated observations that are in the higher end of the propensity score, meaning closer to one, but this

does not account for how much larger the untreated group is. This become more apparent when we look at it after matching where the distribution of the two groups is almost identical.

Overall, I feel comfortable saying that all the matching procedures achieved a satisfactory level of balance, and the results of each estimation can be taken into account when discussing the Collective Bargaining Agreement premium. Before we move on to discussing the estimates from matching, I will briefly explore the drivers of a worker being covered by a Collective Bargaining Agreement, based on the calculated propensity scores.

3.4 Drivers of Collective Bargaining

Previously I already talked about some of the aspects of unionization and through that collective bargaining that we understand, or at least think we understand. Here I will not give an exhaustive discussion of said results, nor will I present the detailed results of the propensity score estimation, as fundamentally this is not the central question of the thesis. I simply want to identify the main drivers of having a Collective Bargaining Agreement and see whether it is consistent with the literature.

Overall, the results back up both what was discussed in the literature review, and in a way, the results of the linear regression. Individual level variables, be it demographic, like age or gender, or job specific, like period of service or employment type, have little to do with the presence of collective bargaining. Rather it was the firm level variables that had a huge effect.

The economy branch variable showed a lot of variance, industries that on average require more physical work and have had historically strong unions in Hungary, like the energy or chemicals sectors, increased the likelihood of having collective bargaining, while those that are more associated with the service sector or seasonal employment, like the IT or hospitality sectors, heavily decrease said likelihood. As also expected, with the increase in the number of

employees and thus the size of a firm, the Collective Bargaining Agreements are more likely to be found.

3.5 Results of the Matching Estimations

As I said before the estimates from matching are the ATT, while the previously discussed estimates are the relative wage gap so they cannot be directly compared. Same goes for all the other estimations that were done for Hungary given that they estimated the same relative wage gap. There are no other papers that used matching to identify the collective bargaining premium in Hungary, thus the only comparison I could make is with papers that used said method for the estimation of the premium in other countries. Going beyond the high-level cross-country comparisons that were done in the literature review would not help in interpreting these specific results, so I am going to restrain from that.

Below, Table 3.5 shows the results of the matching estimations.

	Collective Bargaining Premium estimate	Std. error	95% confidence interval	99% confidence interval
One-to-one MDM	0.0211	0.0084	.0047 to .0375	-.0005-to .0427
Two-to-one MDM	0.0274	0.0072	.0133 to .0416	.0089 to .0460
Three-to-one MDM	0.0216	0.0068	.0083 to .0349	.0042 to .0390
Four-to.one MDM	0.0243	0.0065	.0115 to .0371	.0075 to .0411
One-to-one CBPS	0.0205	0.0085	.0037 to .0372	-.0015 to .0424
Two-to-one CBPS	0.0186	0.0073	.0043 to .0330	-.0002 to .0375
Three-to-one CCBPS	0.0275	0.0069	.0141 to .0409	.0099 to .0451
Four-to-one CBPS	0.0371	0.0079	.0215 to .0527	.0167 to .0575

Table 3.5 – Results of the MDM and CBPS estimations

Because not every estimation yielded results that were significant at the 99% confidence level, I will discuss the intervals that are significant at 95%. For the MDM procedures we can see that they are quite consistent as they put the collective bargaining wage premium somewhere around 2.5%, where the confidence levels range from as low as .5% to as high as above 4.1%. The two-to-one MDM produced the highest estimate at 2.7%, with 1.3% being the lower bound of the estimate, while 4.1% is the upper bound. On the other end the one-to-one MDM gives us the lowest estimate at 2.1%, with the three-to-one MDM producing virtually the same value, with the former's confidence interval being between .5% to 3.7%.

With regards to the CBPS estimates we see much more variance, which is most evidently shown by the fact that whilst the difference between the highest and lowest MDM estimate is .63 percentage points, the same difference for the CBPS is 1.85 percentage points. One of the reasons for this is the fact that the estimate from the four-to-one CBPS matching can be a considered an outlier, since it is one whole percentage point from the nearest estimate, which is the three-to-one CBPS: Nevertheless, the first three CBPS estimates are very close to the MDM ones. The one-to-one and two-to-one CBPS procedures are the most alike, having collective bargaining premium estimates around 2%, with the lower bound being around .4% and the upper bound being around 3.5%. The three-to-one estimate is not far off at 2.75%, with the confidence interval ranging from 1.4% to 4%. As I mentioned before, with its estimate of 3.7%, the four-to-one CBPS estimate is an outlier, and further investigation might be needed to understand the cause of this rather large estimate.

Overall, we have seen an estimate that is slightly higher than the one we saw from the linear regression and is rather low in an absolute sense. Furthermore as the number of matches grew, the quality of the matching process slightly deteriorated, while all of them produced significant results, so with a dataset such as the one that was used here, one-to-one matching seems to do the job. Obviously, I will not be able to determine the amount of bias that endogeneity causes, and this will remain so until such research is recreated with panel data. As far as drawing conclusions from the findings, I will do that in the discussion part.

3.6 Discussion

From all of what was discussed above it is not unreasonable to claim that the collective bargaining wage premium stands somewhere between 1.5 and 2.2 percent, although we have to keep in mind that some upwards bias resulting from endogeneity is present in both estimations. Comparing these estimates to those of previous papers that dealt with the same subject does

offer some perspective, but still these numbers are rather abstract and hard to make sense of. To circumvent this problem, I will try to offer some clarity through a few examples.

The median gross monthly salary for a Hungarian worker in 2019 was 292,320 HUF (hvg.hu, 2021), which adds up to around 3.5 million HUF in a year. This means that said worker takes home about 194,000 HUF every month, or around 2.4 million HUF in a year. Using the more conservative estimate of 1.5 percent, a worker covered by a Collective Bargaining Agreement takes home about 35,000 HUF more on a yearly basis than someone who is not covered by one. This difference grows slightly to around 51,000 HUF if we use the estimate of 2.2 percent.

If we want to be even more realistic, we should take into account the union dues, the membership fees of a union, as well, which is generally around one percent of the monthly salary in Hungary[1]. Obviously not everyone who is covered by such an agreement is affected by these dues but given that the not-so-subtle goal of this thought experiment, and this thesis in general, is to persuade people to join unions, this is not an unreasonable assumption. Subtracting one percent from our estimates leaves us with a collective bargaining wage premium of between 0.5 and 1.2 percent. The yearly premium using the lower bound is about 12,000 HUF, while in the case of the upper bound, it is around 28,000 HUF.

To put these numbers into perspective, at 4000 HUF/person/meal, 12,000 HUF can get someone dinner for three, with delivery (without tipping), while 28,000 HUF buys tickets for four different movies for the same family of three. If we disregard union dues a worker covered by a collective bargaining agreement can afford a cheap smartphone or a used laptop if he or she saves up their yearly wage premium. As I noted earlier, there is more to consider when joining a union than just money, but given that union members, and especially, union organizers,

[1] Members who pay their union dues as individuals (meaning it is not paid by their employers) may reduce the tax base in their annual personal income tax return by the amount of the union dues in accordance with point I.6 of Annex 3 of the Personal Income Tax Act.

sometimes face notable threats, even losing their jobs in extreme cases, for taking part in the labour movement, it is hard to regard these premia as satisfactory motivators.

Conclusion

This thesis attempted to estimate the size of the collective bargaining wage premium in Hungary. How much higher, if at all, are the salaries of those workers who are covered by a collective bargaining agreement as opposed to those who are not? To find this out I used a large dataset on the Hungarian workforce, alas one that lacked reliable data on collective agreements. This problem was solved by consulting an external dataset, which contained historical data on collective agreements in the country. After combining these two I was left with a sizable employee-employer linked dataset, which is particularly suitable for estimating the wage premium because it enables the researcher to exploit firm level variance.

Two estimation strategies were pursued, linear regression and matching. With regards to the former, the estimation procedure had five different stages where with consequent stages new covariates were introduced. In line with the existing literature, until the introduction of firm level variables the coefficient on the Collective Bargaining Agreement was relatively high, but afterwards it declined remarkably. When it came to the latter, both schools of matching, using propensity scores versus using covariate distances, were used in the estimation.

A particular methodological obstacle that I faced with regard to both estimation strategies was that of endogeneity which in this case was a result of the presence of unobservables. When it comes to linear regression this clearly causes the treatment's coefficient to have (upwards) bias, because the exogeneity assumption is violated. With matching the question is a little trickier because the CIA assumption that is possibly violated is not as strict as the exogeneity assumption but claiming that the estimators that result from matching are not biased is no supported by enough evidence.

The linear regression produced a slightly lower estimate of 1.5%, while the different matching estimators put it slightly higher, between 2% and 2.5%. It should not be overlooked that the

two procedures produce different estimators, while the former gives us the relate wage gap, the latter shows the average treatment on the treated. Nevertheless, these estimates can only be described as being rather low and compared to previous research they show a downwards trend.

Before I turn to describing the limitations of the thesis and outlining possible venues of future research, I should note one thing. It is important to keep in mind that the collective bargaining wage premium is just one part of the work unions do, but the magnitude of the wage premium they can achieve has an effect on how successfully they can protect the interests of workers and how organized labour fares at the political level, and vice versa. If we want to understand fully why trade unions are on the path they are on, we must look at the whole picture.

Limitations

There are two major limitations that need to be discussed when it comes to this thesis. One of the I already touched on in this conclusion, endogeneity. Methodologies that can circumvent this problem were either impossible to implement in these circumstances or I was not convinced of their expedience for this case. Thus, whenever the results of the estimations are discussed, we will always have to keep in mind the likely bias that the estimates contain.

The other limitation came in the form of unreliable data. On the one hand, the micro dataset provided by KSH was deemed unfit for research by the data provider. This, of course, was not the fault of KSH, not providing information on Collective Bargaining Agreements, or providing false information on them, has no legal consequences. The database on collective bargaining agreements that was consulted to fill in the gaps that were left by the above issues also was not properly updated, once again because of the lack of legal consequences for non-compliance, thus, expert judgment was needed for the sake of assembling a more reliable dataset. Both of these issues can be traced back to the same problem, the state has no interest in collecting and maintaining high quality data on Collective Bargaining Agreements. If the need for a thorough

understanding of this topic ever arises in Hungary, these institutional problems will need to be addressed.

Future Research

There are two main areas future research will have to address that will greatly improve the accuracy and added value of the thesis. First, something that could have been done during the present iteration of the research, is exploiting the heterogeneity of the collective bargaining wage premium in different industries, or in this particular case, in different branches of the economy. Estimating the desired effect for different branches does not pose a significant hurdle but gives much useful information.

The other area concerns the data that is used in the research. Specifically, presuming that the financial limitations are overcome, data that encompasses a longer time horizon needs to be utilized. This would ensure that the estimated effect is not just an outlier but fits in to the long-term trend of the wage premium. Furthermore, this would enable the usage of methodologies that could potentially solve the thesis' problem with regards to endogeneity.

Appendix A – Categorization of the utilized variables

Gender:

1. Male
2. Female

Highest educational attainment:

10. 0 to 7 years of elementary school
20. 8 years of elementary school
31. Vocational qualification without a high school diploma (e.g., vocational training, vocational school certificate)
32. High school diploma without vocational qualification
33. High school diploma with vocational qualification
40. Secondary vocational qualification certificate obtained in a school system based on graduation
50. Diploma in higher education (tertiary) vocational training
60. Bachelor's degree (BA / BSc) (or equivalent)
70. Master's degree (MA / MSc) or full-time diploma (or equivalent)
80. Diploma (PhD, DLA)

Form of employment:

1. Full-time employees without public employees
2. Part-time, without public employees (average of at least 60 hours worked per month)
3. Employed full time in public employment
4. Employed part-time in public employment
5. According to the employment contract, the employee worked less than 60 hours a month on average

Type of employment contract:

1. Permanent contract
2. Fixed-term contract
3. Student contract
4. Other agreement (e.g., owner, member without employment contract)

Appendix B – Balance Tables

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0064	-
Male	0.0064	-
Permanent employment	0.0000	-
Fixed-term employment	0.0000	-
Full-time contract	0.0078	-
Part-time contract	0.0078	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.0623	1.2491
Econ branch: Manufacture of food products, beverages, and tobacco	0	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0	-
Econ branch: Wood processing, manufacture of paper products, printing	0	-
Econ branch: Manufacture of chemicals and chemical products	0	-
Econ branch: Manufacture of pharmaceuticals	0	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0	-
Econ branch: Manufacture of basic metals and fabricated metal products	0	-
Econ branch: Manufacture of computers, electronic and optical products	0	-
Econ branch: Manufacture of electrical equipment	0	-
Econ branch: Manufacture of machinery	0	-
Econ branch: Manufacture of motor vehicles	0	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0	-
Econ branch: Construction	0	-
Econ branch: Trade, vehicle repair	0	-
Econ branch: Logistics	0.0003	-
Econ branch: Accommodation services, hospitality	0	-
Econ branch: Information, communication	0	-
Econ branch: Financial and insurance activities	0	-
Econ branch: Real estate	0	-
Econ branch: Professional, scientific, and technical activities	0	-
Econ branch: Administrative and support service activities	0	-
Econ branch: Administration, defence; compulsory social security	0	-

Econ branch: Education	0	-
Econ branch: Healthcare	0	-
Econ branch: Social work	0	-
Econ branch: Arts, entertainment, leisure	0	-
Econ branch: Other	0	-
Number of employees	0.1365	1.0283

Table 0.1 - Balance table of the one-to-one MDM

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0037	-
Male	0.0037	-
Permanent employment	0.0000	-
Fixed-term employment	0.0000	-
Full-time contract	0.0078	-
Part-time contract	0.0078	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.0744	1.2982
Econ branch: Manufacture of food products, beverages, and tobacco	0	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0	-
Econ branch: Wood processing, manufacture of paper products, printing	0	-
Econ branch: Manufacture of chemicals and chemical products	0	-
Econ branch: Manufacture of pharmaceuticals	0	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0	-
Econ branch: Manufacture of basic metals and fabricated metal products	0	-
Econ branch: Manufacture of computers, electronic and optical products	0	-
Econ branch: Manufacture of electrical equipment	0	-
Econ branch: Manufacture of machinery	0	-
Econ branch: Manufacture of motor vehicles	0	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0	-
Econ branch: Construction	0	-
Econ branch: Trade, vehicle repair	0	-
Econ branch: Logistics	0.0002	-
Econ branch: Accommodation services, hospitality	0	-
Econ branch: Information, communication	0	-
Econ branch: Financial and insurance activities	0	-

Econ branch: Real estate	0	-
Econ branch: Professional, scientific, and technical activities	0	-
Econ branch: Administrative and support service activities	0	-
Econ branch: Administration, defence; compulsory social security	0	-
Econ branch: Education	0.0009	-
Econ branch: Healthcare	0	-
Econ branch: Social work	0.0007	-
Econ branch: Arts, entertainment, leisure	0	-
Econ branch: Other	0	-
Number of employees	0.1569	1.0129

Table 0.2 - Balance table of the two-to-one MDM

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0027	-
Male	0.0027	-
Permanent employment	0.0000	-
Fixed-term employment	0.0000	-
Full-time contract	0.0113	-
Part-time contract	0.0113	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.0887	1.3478
Econ branch: Manufacture of food products, beverages, and tobacco	0	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0	-
Econ branch: Wood processing, manufacture of paper products, printing	0	-
Econ branch: Manufacture of chemicals and chemical products	0	-
Econ branch: Manufacture of pharmaceuticals	0	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0	-
Econ branch: Manufacture of basic metals and fabricated metal products	0	-
Econ branch: Manufacture of computers, electronic and optical products	0	-
Econ branch: Manufacture of electrical equipment	0	-
Econ branch: Manufacture of machinery	0	-
Econ branch: Manufacture of motor vehicles	0	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0	-
Econ branch: Construction	0.0002	-

Econ branch: Trade, vehicle repair	0	-
Econ branch: Logistics	0.0001	-
Econ branch: Accommodation services, hospitality	0	-
Econ branch: Information, communication	0.0011	-
Econ branch: Financial and insurance activities	0	-
Econ branch: Real estate	0	-
Econ branch: Professional, scientific, and technical activities	0	-
Econ branch: Administrative and support service activities	0.0002	-
Econ branch: Administration, defence; compulsory social security	0	-
Econ branch: Education	0.0023	-
Econ branch: Healthcare	0	-
Econ branch: Social work	0.0009	-
Econ branch: Arts, entertainment, leisure	0	-
Econ branch: Other	0	-
Number of employees	0.1709	1.0092

Table 0.3 - Balance table of the three-to-one MDM

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0049	-
Male	0.0049	-
Permanent employment	0.0000	-
Fixed-term employment	0.0000	-
Full-time contract	0.0132	-
Part-time contract	0.0132	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.0982	1.3794
Econ branch: Manufacture of food products, beverages, and tobacco	0	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0	-
Econ branch: Wood processing, manufacture of paper products, printing	0	-
Econ branch: Manufacture of chemicals and chemical products	0	-
Econ branch: Manufacture of pharmaceuticals	0	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0.0007	-
Econ branch: Manufacture of basic metals and fabricated metal products	0	-
Econ branch: Manufacture of computers, electronic and optical products	0	-
Econ branch: Manufacture of electrical equipment	0	-
Econ branch: Manufacture of machinery	0	-
Econ branch: Manufacture of motor vehicles	0	-
Econ branch: Other manufacturing; installation and	0	-

repair of industrial machinery and equipment		
Econ branch: Electricity, gas, steam, and air conditioning supply	0	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0	-
Econ branch: Construction	0.0002	-
Econ branch: Trade, vehicle repair	0	-
Econ branch: Logistics	0.0003	-
Econ branch: Accommodation services, hospitality	0	-
Econ branch: Information, communication	0.0017	-
Econ branch: Financial and insurance activities	0.0002	-
Econ branch: Real estate	0	-
Econ branch: Professional, scientific, and technical activities	0	-
Econ branch: Administrative and support service activities	0.0002	-
Econ branch: Administration, defence; compulsory social security	0	-
Econ branch: Education	0.0034	-
Econ branch: Healthcare	0	-
Econ branch: Social work	0.0011	-
Econ branch: Arts, entertainment, leisure	0	-
Econ branch: Other	0	-
Number of employees	0.1803	1.0046

Table 0.4 - Balance table of the four-to-one MDM

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0602	-
Male	0.0602	-
Permanent employment	0.0211	-
Fixed-term employment	0.0206	-
Full-time contract	0.0261	-
Part-time contract	0.0261	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.1277	0.8417
Econ branch: Manufacture of food products, beverages, and tobacco	0.0402	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0.0242	-
Econ branch: Wood processing, manufacture of paper products, printing	0.0236	-
Econ branch: Manufacture of chemicals and chemical products	0.0236	-
Econ branch: Manufacture of pharmaceuticals	0.0568	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0.0381	-
Econ branch: Manufacture of basic metals and fabricated metal products	0.0469	-

Econ branch: Manufacture of computers, electronic and optical products	0.0430	-
Econ branch: Manufacture of electrical equipment	0.0558	-
Econ branch: Manufacture of machinery	0.0315	-
Econ branch: Manufacture of motor vehicles	0.0290	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0.0008	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0.2911	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0.0256	-
Econ branch: Construction	0.0007	-
Econ branch: Trade, vehicle repair	0.0308	-
Econ branch: Logistics	0.0549	-
Econ branch: Accommodation services, hospitality	0.0462	-
Econ branch: Information, communication	0.0102	-
Econ branch: Financial and insurance activities	0.0043	-
Econ branch: Real estate	0.0172	-
Econ branch: Professional, scientific, and technical activities	0.0233	-
Econ branch: Administrative and support service activities	0.0285	-
Econ branch: Administration, defence; compulsory social security	0.0477	-
Econ branch: Education	0.0168	-
Econ branch: Healthcare	0.0328	-
Econ branch: Social work	0.0146	-
Econ branch: Arts, entertainment, leisure	0.0356	-
Econ branch: Other	0.0712	-
Number of employees	0.1017	1.0026

Table 0.5 – Balance table of the one-to-one CBPS

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0666	-
Male	0.0666	-
Permanent employment	0.0147	-
Fixed-term employment	0.0143	-
Full-time contract	0.0389	-
Part-time contract	0.0389	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.0988	0.9147
Econ branch: Manufacture of food products, beverages, and tobacco	0.0356	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0.0290	-
Econ branch: Wood processing, manufacture of paper products, printing	0.0254	-

Econ branch: Manufacture of chemicals and chemical products	0.0226	-
Econ branch: Manufacture of pharmaceuticals	0.0735	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0.0396	-
Econ branch: Manufacture of basic metals and fabricated metal products	0.0456	-
Econ branch: Manufacture of computers, electronic and optical products	0.0299	-
Econ branch: Manufacture of electrical equipment	0.0536	-
Econ branch: Manufacture of machinery	0.0402	-
Econ branch: Manufacture of motor vehicles	0.0502	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0.0081	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0.3068	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0.0289	-
Econ branch: Construction	0.0114	-
Econ branch: Trade, vehicle repair	0.0433	-
Econ branch: Logistics	0.0657	-
Econ branch: Accommodation services, hospitality	0.0352	-
Econ branch: Information, communication	0.0143	-
Econ branch: Financial and insurance activities	0.0101	-
Econ branch: Real estate	0.0101	-
Econ branch: Professional, scientific, and technical activities	0.0613	-
Econ branch: Administrative and support service activities	0.0285	-
Econ branch: Administration, defence; compulsory social security	0.0261	-
Econ branch: Education	0.0541	-
Econ branch: Healthcare	0.0056	-
Econ branch: Social work	0.0342	-
Econ branch: Arts, entertainment, leisure	0.0146	-
Econ branch: Other	0.0282	-
Number of employees	0.1042	0.9964

Table 0.6 – Balance table of the two-to-one CBPS

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0326	-
Male	0.0326	-
Permanent employment	0.0139	-
Fixed-term employment	0.0134	-
Full-time contract	0.0313	-
Part-time contract	0.0313	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.1028	0.9002

Econ branch: Manufacture of food products, beverages, and tobacco	0.0413	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0.0330	-
Econ branch: Wood processing, manufacture of paper products, printing	0.0202	-
Econ branch: Manufacture of chemicals and chemical products	0.0284	-
Econ branch: Manufacture of pharmaceuticals	0.0806	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0.0310	-
Econ branch: Manufacture of basic metals and fabricated metal products	0.0420	-
Econ branch: Manufacture of computers, electronic and optical products	0.0368	-
Econ branch: Manufacture of electrical equipment	0.0546	-
Econ branch: Manufacture of machinery	0.0389	-
Econ branch: Manufacture of motor vehicles	0.0453	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0.0151	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0.2258	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0.0940	-
Econ branch: Construction	0.0085	-
Econ branch: Trade, vehicle repair	0.0394	-
Econ branch: Logistics	0.0604	-
Econ branch: Accommodation services, hospitality	0.0308	-
Econ branch: Information, communication	0.0108	-
Econ branch: Financial and insurance activities	0.0115	-
Econ branch: Real estate	0.0090	-
Econ branch: Professional, scientific, and technical activities	0.0192	-
Econ branch: Administrative and support service activities	0.0217	-
Econ branch: Administration, defence; compulsory social security	0.0278	-
Econ branch: Education	0.0482	-
Econ branch: Healthcare	0.0042	-
Econ branch: Social work	0.0375	-
Econ branch: Arts, entertainment, leisure	0.0155	-
Econ branch: Other	0.0247	-
Number of employees	0.1019	0.9967

Table 0.7 – Balance table of the three-to-one CBPS

	Standardized mean difference	Variance ratio (if applicable)
Female	0.0437	-

Male	0.0437	-
Permanent employment	0.0159	-
Fixed-term employment	0.0157	-
Full-time contract	0.0349	-
Part-time contract	0.0349	-
Subsidized full-time	0	-
Subsidized part-time	0	-
Period of service	0.0874	0.9430
Econ branch: Manufacture of food products, beverages, and tobacco	0.0392	-
Econ branch: Manufacture of textiles, clothing, leather, and leather products	0.0342	-
Econ branch: Wood processing, manufacture of paper products, printing	0.0229	-
Econ branch: Manufacture of chemicals and chemical products	0.0252	-
Econ branch: Manufacture of pharmaceuticals	0.0890	-
Econ branch: Manufacture of rubber, plastic, and non-metallic mineral products	0.0237	-
Econ branch: Manufacture of basic metals and fabricated metal products	0.0357	-
Econ branch: Manufacture of computers, electronic and optical products	0.0384	-
Econ branch: Manufacture of electrical equipment	0.0450	-
Econ branch: Manufacture of machinery	0.0391	-
Econ branch: Manufacture of motor vehicles	0.0486	-
Econ branch: Other manufacturing; installation and repair of industrial machinery and equipment	0.0178	-
Econ branch: Electricity, gas, steam, and air conditioning supply	0.2536	-
Econ branch: Water supply; wastewater collection and treatment, waste management, decontamination	0.0677	-
Econ branch: Construction	0.0141	-
Econ branch: Trade, vehicle repair	0.0344	-
Econ branch: Logistics	0.0657	-
Econ branch: Accommodation services, hospitality	0.0331	-
Econ branch: Information, communication	0.0156	-
Econ branch: Financial and insurance activities	0.0238	-
Econ branch: Real estate	0.0110	-
Econ branch: Professional, scientific, and technical activities	0.0198	-
Econ branch: Administrative and support service activities	0.0204	-
Econ branch: Administration, defence; compulsory social security	0.0293	-
Econ branch: Education	0.0530	-
Econ branch: Healthcare	0.0028	-
Econ branch: Social work	0.0416	-
Econ branch: Arts, entertainment, leisure	0.0143	-
Econ branch: Other	0.0212	-

Number of employees	0.1022	0.9921

Table 0.8 - Balance table of the four-to-one CBPS

Appendix C – Distribution of propensity scores in the different

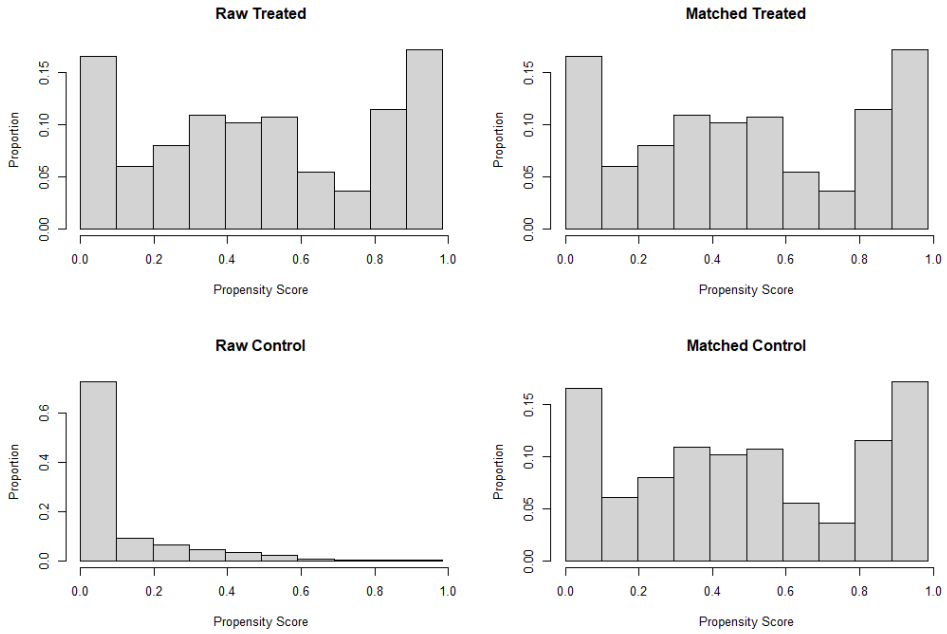


Figure 0.1 - Distribution of propensity scores for one-to-one CBPS

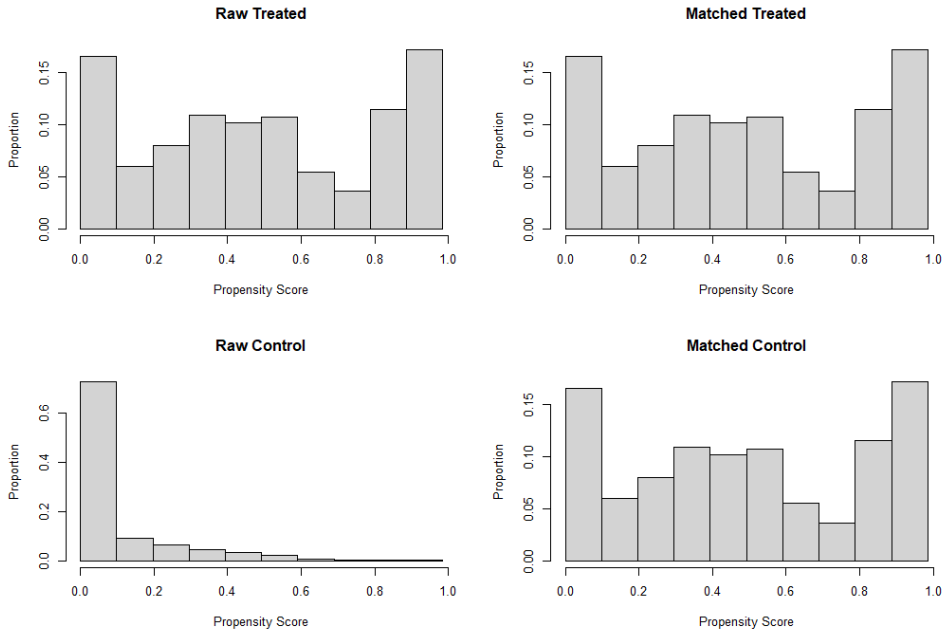


Figure 0.2 - Distribution of propensity scores for two-to-one CBPS

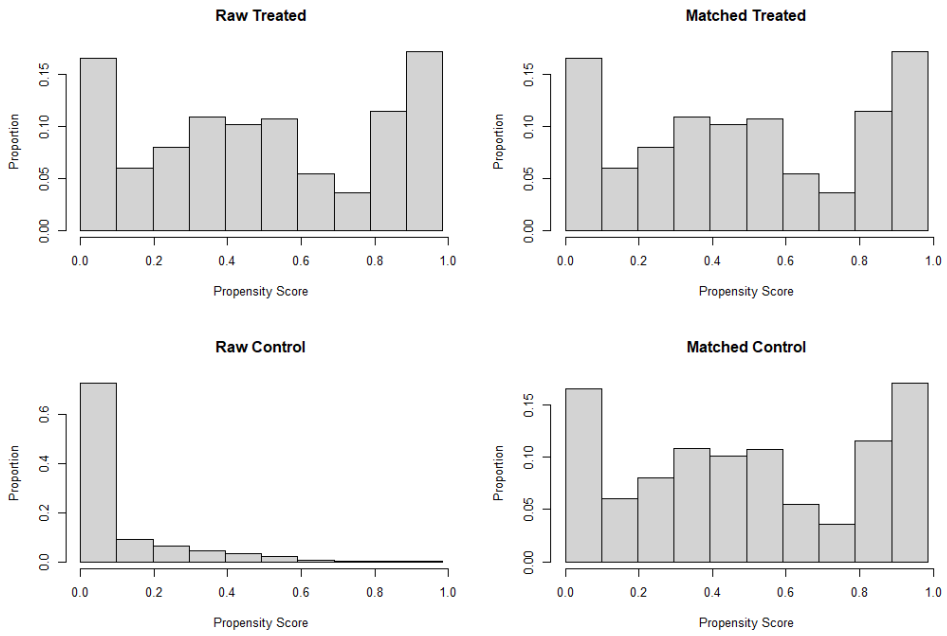


Figure 0.3 - Distribution of propensity scores for three-to-one CBPS

Appendix D- Supplementary materials

Script for the Statistical Analysis conducted in R Studio, uploaded to GitHub:

https://github.com/kisgutzi/thesis/blob/main/r%20file%20gutzianas%20ioannis.R

Bibliography

Fang, T., & Verma, A. (2002). Union Wage Premium. *Perspectives on Labour and Income 14(9)*, 13-19.

Antonczyk, D. (2011). *Using social norms to estimate the effect of collective bargaining on the wage structure.* Job Marker Paper.

Ashenfelter, O., Layard, R., & Card, D. E. (1986). *Handbook of Labor Economics, Vol II.* New York: Elsevier.

Ashraf, J. (1992). Union wage premiums in an instrumental variables framework. *Journal of Labor Research, volume 13*, 231-236.

Austin, P. (2009). Balance Diagnostics for Comparing the Distribution of Baseline Covariates Between Treatment Groups in Propensity-Score Matched Samples. *Statistics in Medicine 28 (25)*, 3083–3107.

Balczer, B. (2016, 12 19). Is it allowed to strike in Hungary? (In Hungarian: Szabad-e sztrájkolni Magyarországon?). *Új Egyenlőség*.

Berke, G. (2015). The Collective Agreement in the New Labour Code. In G. Kiss, *Trade unions and collective agreements in the new Labour Code.*

Bíró Nagy, A., Kadlót, T., & Varga, Á. (2012). *Job creation in Hungary (In Hungarian: Munkahelyteremtés Magyarországon).* Policy Solutions.

Blackburn, M. L. (2008). Are Union Wage Differentials in the United States Falling? *Industrial Relations, Volume 47, Issue 3*, 390-418.

Blanchflower, D., & Bryson , A. (2004). What Effect Do Unions Have on Wages Now and Would Freeman and Medoff Be Surprised? *Journal of Labor Research 25(3)*, 383-414.

Blanchflower, D., & Bryson , A. (2007). *The Wage Impact of Trade Unions in the UK Public and Private Sectors.* IZA DP No. 3055.

Booth, A., & Bryan, M. (2004). The Union Membership Wage-Premium Puzzle: Is There a Free Rider Problem? *ILR Review Vol. 57, No. 3*, 402-421.

Breda, T. (2014). Firms' Rents, Workers' Bargaining Power and the Union Wage Premium. *The Economic Journal*, 1616-1652.

Bronars, S., & Lot, J. (1989). Why Do Workers Join Unions? The Importance Of Rent Seeking. *Economic Inquiry Volume 27, Issue 2*, 305-324.

Bryson, A. (2002). *The Union Membership Wage Premium: An Analysis Using Propensity Score Matching.* CEP Discussion Papers 530.

Caliendo, M., & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys 22 (1)*, 31–72.

Calmfors, L., & Driffill, J. (1988). Bargaining Structure, Corporatism and Macroeconomic Performance. *Economic Policy Vol. 3, No. 6*, 13-61.

Campolieti, M. (2018). Matching and Inverse Propensity Weighting Estimates of the Union Wage Premium: Evidence from Canada, 1997–2014. *Industrial Relations, Vol. 57, No. 1* , 101-130.

Card, D., & De La Rica, S. (2006). Firm-Level Contracting and the Structure of Wages in Spain. *ILR Review, Vol. 59, Issue 4*, 573-592.

Choi, H., & Ramos, R. (2021). The union wage premium in a segmented labor market: new evidence from Korea. *Journal of the Asia Pacific Economy*.

Congleton, R., Hilman, A., & Konrad, K. (2008). *40 Years of Research on Rent Seeking 2.* Heidelberg: Springer Berlin.

Conolly, R., Hirsch, B., & Hirschey, M. (1986). Union Rent Seeking, Intangible Capital, and Market Value of the Firm. *The Review of Economics and Statistics Vol. 68, No. 4*, 567-577.

Crockett, G., & Hall , K. (1986). *Why Do Salaried Professionals Join Unions: An Australian Perspective.* WA Labour Market Research Centre Discussion Paper No. 14.

Cunningham , S. (2021). *Causal Inference: The Mixtape.* Yale University Press.

Diamond , A., & Sekhon, J. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies . *The Review of Economics and Statistics 95 (3)*, 932–945.

DiNardo, J., & Lee, D. (2004). Economic Impacts of New Unionization on Private Sector Employers: 1984–2001. *The Quarterly Journal of Economics, Vol. 119, Issue 4*, 1383–1441.

Duncan, G., & Stafford, F. (1980). Do union members receive compensating wage differentials,. *American Economic Review, 70 (3)*, 355-371.

Éber, M. (2020). *The drop (In Hungarian: A csepp).* Budapest: Napvilág.

Eren, O. (2007). Measuring the Union–Nonunion Wage Gap Using Propensity Score Matching. *Industrial Relation, Vol. 46, No. 4*, 766-780.

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism.* Cambridge: Polity Press.

ETUI. (2019, 04 08). *Industrial relations in Hungary: background summary (update March 2019)*. Retrieved from etui.org: https://www.etui.org/covid-social-impact/hungary/industrial-relations-in-hungary-background-summary-update-march-2019

Farber, H. (1986). The analysis of Union Behavior. In O. Ashenfelter, R. Layard, & D. Card, *The Handbook of Labor Economics* (pp. 1039-1089). New York: Elsevier.

Fong, C., Ratkovic, M., & Imai, K. (2021). *CBPS: Covariate Balancing Propensity Score.* https://CRAN.R-project.org/package=CBPS.

Frandsen, B. (2012). Why Unions Still Matter: The Eects of Unionization on the Distribution of Employee Earnings.

Freeman, R., & Medoff , J. (1984). *What Do Unions Do?* New York: Basic Books.

Greifer, N. (2022, 05 21). *Matching Methods*. Retrieved from https://cran.r-project.org/: https://cran.r-project.org/web/packages/MatchIt/vignettes/matching-methods.htm

Grindt, R. (2013). *Trade unions in Hungary 2012: New hope in a year of woes?* Friedrich Ebert Stiftung .

Gürtzgen, N. (2016). Estimating the wage premium of collective wage contracts: Evidence from longitudinal linked employer-employee data. *Industrial Relations: A Journal of Economy and Society*, 294-322.

Gyulavári, T., & Kártyás, G. (2015). *Th Hungarian Flexicurity Pathway: New Labour Code after Twenty Years in the Market Economy.* Budapest: Pázmány Péter Catholic University.

Hainmuller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis 20*, 25-46.

Hajdú, J. (2019). Hungary. In F. Hendrickx, *IEL Labour Law* (pp. 1-311). Kluwer Law International BV.

Hartog, J., Leuven, E., & Teulings, C. (2002). Wages and the bargaining regime in a corporatist setting: the Netherlands. *European Journal of Political Economy, Vol. 18, Issue 2*, 317-331.

Hatos, P. (2018). *The Cursed Republic - The History of the 1918 Collapse and Revolution (In Hungarian: Az elátkozott köztársaság - Az 1918-as összeomlás és forradalom története).* Budapest: Jaffa.

Heckman, J., & Navarro, D. (2004). Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choices. *Review of Economics and Statistics 86(1)*, 30-57.

Heckman, J., Hidehiko, I., Smith, J., & Petra, T. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica 66(5)*, 1017-1098.

Héthy, L. (1994). Political Changes and the Transformation of Industrial Relatíons in Hungary. In J. Rilan, R. Lansbury, & C. Verevis, *The Future of Industrial Relations. Global Change and Challange.* London: Sage.

Hirano, K., Imbens, G., & Ridder, G. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica 71 (4)*, 1161–1189.

Hirsch, B. (2004). Reconsidering union wage effects: Surveying new evidence on an old topic. *Journal of Labor Research, volume 25*, 233–266 .

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables.* https://CRAN.R-project.org/package=stargazer.

Ho, D., Imai , K., King, G., & Stuart , E. (2011). MatchIt: Nonparametric Preprocessing for Parametric. *Journal of Statistical Software, Vol. 42, No. 8*, 1-28.

Humphreys, M. (2009). *Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities.* New York: Columbia University.

Huntington-Klein, N. (2021). *The Effect: An Introduction to Research Design and Causality.* CRC Press.

hvg.hu. (2021, 02 23). 75 ezer forinttal alacsonyabb a mediánbér az átlagnál. *HVG*.

ILO. (2022, 05 14). *Industrial relations*. Retrieved from ilo.org: https://www.ilo.org/ilostat-files/Documents/description_IR_EN.pdf

Imai, K., & Ratkovic, M. (2014). Covariate Balancing Propensity Score. *Journal of the Royal Statistical Series B 76(2)*, 243-263.

Iversen, T., Pontusson, J., & Soskice, D. (2000). *Unions, employers, and central banks: Macroeconomic coordination and institutional change in social market economies.* Cambridge: Cambridge University Press.

Kertesi, G., & Köllő János. (2003b). Industrial Wage Differences in Hungary, Rent bargaining in concentrated industries in the presence of trade unions (In Hungarian: Ágazati bérkülönbségek Magyarországon 2. rész: Járadékokon való osztozkodás koncentrált ágazatokban, szakszervezeti aktívitás. *Közgazdasági Szemle*, 1049-1074.

Kertesi, G., & Köllő, J. (2003a). Industrial Wage Differences in Hungary (In Hungarian: Ágazati bérkülönbségek Magyarországon 1. rész: Az ágazati járadékképződés alternatív modelljei). *Közgazdasági Szemle*, 923-938.

Kertész, G. (2020). "Slave Act" - Legal Analysis and Perspectives (In Hungarian: "Rabszolgatörvény" - Jogi elemzés és perspektívák). *Multidiszciplináris kihívások, sokszínű válaszok - Gazdálkodás- és Szervezéstudományi folyóirat , (1)*, 61-70.

King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis 27 (4)*, 435-454.

Kitschelt, H. (1994). *The Transformation of European Social Democracy*. New York: Cambridge University Press.

Kohl, H. (2015). Convergence and divergence – 10 years since EU enlargement. *Transfer: European Review of Labour and Research*, 285-311.

KSH. (2021). *Sztárjkok*. Retrieved from ksh.hu: https://www.ksh.hu/stadat_files/mun/hu/mun0073.html

Kun, A. (2019). Wage-setting in Hungary - From a Labour Law Perspective. *Italian Labour Law e-Journal Issue 2, Vol. 12*, 23-43.

Lee, D., & Mas, A. (2012). Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961–1999. *The Quarterly Journal of Economics, Volume 127, Issue 1*, 333-378.

Lewis, G. (1986). Union relative wage effect. In O. Ashenfelter, R. Layard, & D. E. Card, *Handbook of Labor Economics, Vol II* (pp. 1139-1181). New York: Elsevier.

Lux, J. (2008). *The History of Trade Unions in Hungary (In Hungarian: A magyarországi szakszervezetek történetéből).* Budapest: Friedrich Ebert Stiftung.

Mac Flynn, P. (2020). *The Impact of Collective Bargaining on pay in Northern Ireland.* NERI Working Paper Series No. 66.

Mahony, C. (2014, 02 11). *Effects of dimensionality on distance and probability density in climate space*. Retrieved from The Season Alter: http://blogs.ubc.ca/colinmahony/2014/02/11/dimensionalityeffects/

MOSZ. (2022). *Bemutatkozás*. Retrieved from munkastanacsok.hu: https://munkastanacsok.hu/rolunk/bemutatkozas/

Nahm, D., Dobbie, M., & MacMillan, C. (2017). Union wage effects in Australia: an endogenous switching approach. *Applied Economics, 49:39*, 3927-3942.

Neumann, L. (1997). Company Restructuring and Industrial Relations in Privatised Telecommunication Services in Hungary. *Paper presented to the Fifth IIRA European Regional Congress.* Dublin.

Neumann, L. (2001). Do Decentralized Collective Bargaining Have an Impact on the Labour Market in Hungary? *European Journal of Industrial Relations, Vol. 8, No. 1*, 11-31.

Neumann, L. (2006). *Collective Agreements – Still Decentralized, with Shrinking Coverage.* The Hungarian Labour Market, Review and Analysis.

OECD. (2022, 05 21). *stats.oecd.org*. Retrieved from OECD.Stat: https://stats.oecd.org/index.aspx?DataSetCode=CBC

R Core Team. (2020). *R: A language and environment for.* Vienna: R Foundation for Statistical.

Renaud , S. (1998). Unions, Wages and Total Compensation in Canada. *Industrial Relations, Vol. 53, No. 4*, 710-729.

Rigó, M. (2008). *Estimating Union – Non-union Wage Differential in Hungary.* Central European University.

Robinson, C. (1989). The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Methods. *Journal of Political Economy 97*, 639-667.

Rubin, D. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, 20-36.

Scheiring, G. (2019). *The death of a democracy (In Hungarian: Egy dmeokrácia halála: Az autoriter kapitalizmus és a felhalmozó állam felemelkedése Magyarországon).* Budapest: Napvilág.

Sebők, M. (2019). *In the captivity of paradigms: Elites and ideologies in Hungarian financial capitalism (In Hungarian: Paradigmák fogságában: Elitek és ideológiák a magyar pénzügyi kapitalizmusban).* Budapest: Napvilág.

Selig, P. (2006). *A History of Trade Unionism in the United States.* New York: Hard Press.

Stephan, G., & Gerlach, K. (2005). Wage settlements and wage setting: results from a multi-level model. *Applied Economics, Col. 37, Issue 20*, 2297-2306.

Tóth, A. (2000). Hungarian unions in the midst of politics and the economy (In Hungarian: A magyar szakszervezetek a politika és a gazdaság viharába). *Politikatudományi Szemle*, 151-184.

Upchurch, M., Taylor Graham, & Mathers, A. (2009). *The Crisis of Social Democratic Trade Unionism in Western Europe: The Search for Alternatives.* Aldershot: Ashgate.

Wittenberg, M., & Kerr, A. (2019). *Wage determination in South Africa: The case of the union wage premium since the end of apartheid.* REDI3x3 Working paper 56.