STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's Thesis in Economics
Academic Year 2021-22

Does Digitization Impact Job Matching Efficiency? A Study of Swedish Labor Market Regions

Victoria Gruner (41886)

Abstract

The use of digital tools in recruiting has increased over the last 20 years and is likely to have improved the transparency and shortened the duration of job search. Therefore, I asses empirically whether increased digitization in firms has had a positive impact on the efficiency of job matching. I estimate the correlation between firm digitization and matching efficiency within Swedish labor market regions using panel data between the years 1998 and 2016. Firm digitization is approximated by the number of individuals who work in a labor market region and possess formal education in the field of information technology. My results are mixed, while I do find a significant positive correlation in some regressions, robustness tests do not confirm the effect. With the inclusion of year fixed effects the estimated correlation is significantly reduced. Therefore, this study can not confirm that digitization improves job matching efficiency.

Keywords: Job Matching Efficiency, Job Search, Labor Market, Digitization, Sweden
JEL: O330, J630, J640, R230

Supervisors:	Sampreet Goraya and Jonas Kolsrud
Date submitted:	May 16, 2022
Date examined:	May 31, 2022
Discussant:	Andrew Buckner
Examiner:	Kelly Ragan

Acknowledgements

I want to thank my supervisor Jonas Kolsrud at the National Institute of Economic Research (Konjunkturinstitutet) for his valuable feedback at different stages of the research process, for the insightful discussions about the Swedish labor market and for organizing the access to relevant data sources. I am very grateful for the time and thought he dedicated to support this project.

In addition, I want to thank my supervisor at the Stockholm School of Economics, Sampreet Goraya, for his guidance and very thoughtful comments throughout the semester.

Contents

List of Figures

List of Tables

1 Introduction

The way in which people search and apply for jobs has been fundamentally transformed in the last decades. In the 1990s, searching for a job meant reading newspapers or physically going to a job board. The amount of information about the job entailed in a newspaper advertisement was limited, since space was costly. Applications were physically sent out by mail. The firms manually sorted applications and sent a reply to the applicant by mail with an invitation to an interview. Applying for a job in 2022 is quite different. Job seekers might start the job search on their smartphones by scrolling through a professional social network where they get shown job offers matching their skill profile. At the same time their search might be detected by the algorithm, which then suggests them as potential candidates to recruiters on the platform. More information about the job can be easily obtained by visiting the firm website and asking questions to a chat bot. The actual application just entails filling out an online application form. At the firm, these online applications might get sorted and pre-selected by an algorithm, which shortens the time until the applicant gets a reply. Software might be used to automatize the scheduling of an interview. In the further selection, firms might employ tests and online interviews which are automatically assessed without a recruiter (Black and van Esch 2020, Ignatova et al. 2018). This development suggests that job search has become more transparent and shorter in time. The efficiency of job matching is very relevant for the labor market, as it would be predicted to lower the average unemployment duration and unemployment itself. Therefore, I study whether the increased use of digital tools in firms has had a positive effect on job matching efficiency. I use panel data on Swedish labor market regions between 1998 and 2016 and analyze how changes of information technology (IT) use in firms affect the job matching efficiency within these regions. IT use in firms is approximated by the number of individuals with a formal IT education who work in these regions. While most previous research in this area has focused exclusively on the effects of online job search, I contribute to literature because I consider the general effect of digitization on matching efficiency over a longer period of time. Moreover, this is one of the first studies in this regard on the Swedish labor market.

The study is structured as follows: Chapter 2 motivates the research question by elaborating the technological advancements in recruiting and their relevance for matching efficiency according to search theory. The chapter gives furthermore an overview of previous literature in the field. In Chapter 3 I briefly describe my identification strategy and the data I use in the different stages of my analysis. Chapter 4 outlines how I compute matching efficiency for each labor market region

and presents the results. Chapter 5 describes the methodological assumptions for using the approximation variable, as well as the actual measurement of this proxy variable. Finally, in chapter 6, I analyze how digitization is correlated with matching efficiency. The main tools are fixed effects and first difference regressions. Section 6.1 describes the methodology and limitations of this analysis, namely the functional forms of regressions and the included controls. Section 6.2 presents and discusses the results.

2 Theoretical Background

This chapter consists of two parts, in the first part I motivate my research question by reviewing the technological developments in recruiting since the mid 1990s in more detail. Then, I provide some theoretical background on the concept of matching efficiency and explain why recruiting technology might be influential. In the second part I review previous literature on this question.

2.1 Motivation of the Research Question

2.1.1 Technological Advancements in Recruiting

Since the mid-1990s the technology use in recruiting has increased steadily. Until then recruiting was done entirely analogue, vacancies were posted on job boards and advertised in newspapers. In the 1990s the first online job boards emerged e.g. Monster.com. For firms they expanded the pool of applicants to a larger geographic area and, in contrast to newspaper advertisements, allowed firms to include as much information and detail in a job description as they wished without extra costs. Accordingly, for applicants this meant that they could obtain information on vacancies faster than before, because they did not have to go to a job board physically or buy newspapers, and that they could search in a larger geographic area more easily. These online job boards saw high growth between mid-1990s and the mid-2000s. This change has been coined 'The First Digital Revolution' in recruiting. The second one started in the mid-2000s with 1) the emergence of aggregated search engines e.g. Indeed, which simplified online job search even more by searching multiple websites and 2) professional social networks e.g. LinkedIn, which helped companies to target specific job seekers more efficiently and increased the possibilities for professional networking. These platforms became widespread starting in 2010. Overall, the previously mentioned developments lead to an increase in applications with the consequence that HR professionals had to sort a much higher amount of applications. This triggered the third step in the digital acceleration, the use of software

in general and artificial intelligence (AI). These tools are used for i) outreach, ii) for identifying fitting candidates, iii) targeting them with job postings which are individually customized with machine learning methods and iv) facilitating the assessment of the applications. In this regard the main advantage of AI is to not further increase the number of applicants but to increase the fit of the applicants, which in turn saves time for the recruitment team. When it comes to assessing the incoming applications. Software is used for example, to screen CVs, employ tests and games or for assessing video interviews. Software use also improves the ease of an application for the applicant e.g. filling in an online application instead of mailing an application, texting with chat bots to answer questions instead of phoning the firm (Black and van Esch 2020, Hmoud et al. 2019). A LinkedIn report interviewing recruiters world wide in 2017, shows that 76% of respondents across all countries say that AI will have at least a somewhat significant impact. The most important driver identified is, that it saves time (67 % agree), it removes human bias (43%), saves money (30%) and delivers best candidate matches (31%). According to the respondents AI is mostly used for sourcing and for screening candidates (Ignatova et al. 2018). Lisa and Talla Simo (2021) study the implementation of AI in recruiting in Sweden by conducting interviews with recruiting professionals from different sectors. They find that AI is mostly used in the early recruitment stages and confirm large savings in time, improved efficiency and a potential decrease in human biases.

2.1.2 Search Theory and Matching Efficiency

Search theory is a branch of labor economics which aspires to explain why vacancies and unemployment coexist. The permanent baseline level of unemployment is called the Equilibrium Rate of Unemployment. In a frictionless neo-classical Walrasian model of the labor market, unemployment does not exist because labor supply and labor demand meet in equilibrium (Yashiv 2007). On the one hand, the existence of a baseline level of unemployment can be explained by wage rigidities, which prevent wage levels to adjust to the equilibrium level and therefore cause unemployment. These wage rigidities can be caused by various factors, such as collective wage bargaining by labor unions, minimum wages or the generosities of unemployment benefits which raise individual reservation wages (Layard et al. 2005).

While this theory does make a relevant contribution to explain the existence of unemployment, search theory attempts to explain the coexistence of unemployment and vacancies by focusing on non-wage factors. The basic assumption of search theory is that labor markets do not exhibit perfect information. A new job seeker will not immediately obtain all information about all vacancies, furthermore apply-

ing for jobs and getting feedback from the employer takes time. Therefore, search theory introduces the concept of matching efficiency, also called search effectiveness. Matching efficiency is thought to depend on the one hand, on decisions by the job seeker, such as the time and effort spent searching and the probability to accept a job offer, and on the other hand, on the recruiting practices of employers. In this regard technological changes, for example shifting advertisements from newspapers to the internet are considered to increase matching efficiency. Increased funding for government owned employment services could also increase matching efficiency (Petrongolo and Pissarides 2001, Layard et al. 2005).

Search theories usually model a matching process and include matching efficiency as a parameter. A common model for the process of job search and hiring is a matching function, which depicts a trade technology between job seekers (unemployed) and employers producing a number of job seekers who get hired, also called matches. The number of matches is assumed to depend on the number of vacancies, unemployed and on the matching efficiency. The most basic specification of a matching function is

$$M = m(U, V) \tag{1}$$

where M stands for matches, that means the number of unemployed who found a job, U for the number of unemployed and V for the number of vacancies. The function is usually assumed to be increasing in U and V and concave. Commonly the stock of unemployment and the stock of vacancies are used as regressors. Most studies find that a Cobb Douglas function with constant returns to scale matches the empirical data. The following functional form is widespread in the literature

$$M = AU^{\gamma}V^{1-\gamma} \tag{2}$$

Here, A represents the matching efficiency and γ is a scale parameter. Most studies estimate this function as a log linear approximation (Petrongolo and Pissarides 2001, Bouvet 2012). This model would predict, that an increase in matching efficiency leads to a higher number of matches for a given level of unemployment and vacancies.

Another model which is often used in search theory is the Beveridge curve. The Beveridge curve depicts the relationship between vacancies and unemployment, which is predicted to be inverse. That means, given a high level of vacancies, a low level of unemployment is expected and vice versa. To derive this implication, it is assumed that, in steady state, the number of matches is equal to the number of separations. Therefore, in the simple matching function $M = AU^{\gamma}V^{1-\gamma}$, matches M are substituted with separations S. Furthermore, U, V and S are divided by the

labor force and denoted with lower case letters.[1]

$$s = Au^{\gamma}v^{1-\gamma} \tag{3}$$

Rearranging the equation

$$u = (\frac{s}{Av^{1-\gamma}})^{\frac{1}{\gamma}} \tag{4}$$

Because the rate of separations s is assumed to be fixed in steady state, an inverse relationship between unemployment and vacancies is predicted, which constitutes the Beveridge curve. When V increases, U is predicted to decrease according to this equation. Therefore, cyclical shocks to the economy may increase or decrease unemployment and vacancy levels, but they only cause shifts along the curve. A shift of the curve itself must be caused by a change in matching efficiency A. The curve shifts inwards, if matching efficiency increases and outwards, if matching efficiency deteriorates. Therefore, an increase in matching efficiency would cause a decline in the predicted unemployment level for a given number of vacancies (Bouvet 2012).

In this section I have shown that digital technology has changed recruiting practices and job seeking by making information on vacancies and job seekers more transparent and by accelerating the application and recruitment process time wise. The considered technology spans from using internet job search to sophisticated artificial intelligence software. In theory, such changes in recruitment practices are expected to have a positive effect on matching efficiency. An increase in matching efficiency is relevant for the labor market since it would lead to a higher number of job matches and lower unemployment for a given level of vacancies. Therefore, it is interesting to assess empirically whether the increased use of digital technology has indeed lead to an increase in matching efficiency. In the next section I will consider previous research on this question.

2.2 Previous Research

In this part I first consider previous research on the development of matching efficiency in general and second I review research on the relationship between digitization and matching efficiency.

[1] M, S, U and V are normalized to the respective Labor force and then denoted by lower case letters u,v, s and m, $u = \frac{U_t}{L_t}$, $v_t = \frac{V_t}{L_t}$, $s_t = \frac{S_t}{L_t}$ and $m_t = \frac{M_t}{L_t}$ where L_t represents the number of persons in the labor force.

2.2.1 Job Matching Efficiency in Sweden

Empirical evidence suggests that Beveridge curves shift frequently over time, indicating that matching efficiency changes frequently and due to various reasons (Bouvet 2012). The Beveridge curve in Sweden has shifted outwards in the time period 1981 to 2014, indicating a decline of matching efficiency on the Swedish labor market. Especially after the economic crisis 2008/2009 matching deteriorated. While the vacancy rate now even exceeds the pre-crisis levels, the unemployment rate has remained at the higher crisis level (Eklund et al. 2015). The reasons for this shift are not certain, but some events can be suspected to have played a role. In 2006 a new government came into office which reformed the sickness and disability benefit system to decrease the number of beneficiaries. While the goal of the reform was to improve the employment rate among older workers, it is also suspected that it increased unemployment in this group, because older workers have difficulties finding a new job (Spasova et al. (2016), OECD (2009)). Another relevant aspect is the increased inflow of migrants. In 2015 the number of newly arriving migrants in Sweden was the highest number per capita which was ever registered in an OECD country. Integrating migrants into the labor market has been identified as challenging for various reasons. Migrants have on average lower education levels than the Swedish population, at the same time Sweden has a very high share of jobs requiring at least upper secondary education and generally high entry wages (OECD 2016). Another possible reason is a higher mismatch between the qualifications of the unemployed and those required by the employers. This could be caused by individual education decisions or by a failure of the education system to adapt to changes in labor demand (Eklund et al. 2015).

However, in a number of other developed countries, matching efficiency seems to have deteriorated as well. France, the Netherlands, Spain, the US, and Italy experienced similar changes (Bonthuis et al. 2016, Eklund et al. 2015). Considering other countries as well, matching efficiency seems to be also influenced by factors not determined by individual job seekers or firms recruitment practices. Changes in the socioeconomic composition of the unemployment pool are widely considered to explain changes in matching efficiency (Barnichon and Figura 2011). High shares of long-term unemployment decrease matching efficiency, since they cause a decline in human capital and signal low abilities to employers (Bouvet 2012). Higher unemployment benefits and higher minimum wages are associated with lower matching efficiency. Such institutions lead to higher reservation wages, which could cause individuals to be more picky about offered jobs (Bouvet 2012). Furthermore, sectoral and geographic shifts in the economy change the skills required and geographic

distribution of new vacancies and decrease matching efficiency at least temporarily (Petrongolo and Pissarides 2001).

Overall, it can be observed that matching efficiency has not increased during the time period in which the usage of digital tools expanded. This development could be driven by various other factors and does not identify how matching efficiency would have evolved absent the increase in digitization. In the next section I will therefore focus on previous research on the direct effect of digitization on matching efficiency.

2.2.2 Effects of Digitization on Job Matching

The majority of previous research in this field has been focusing on the effects of online job posting and online job search on matching efficiency. One of the first studies was conducted by Kuhn and Skuterud (2004). They used employee data from 1998 and 2000 and found that unemployment spells are shorter for job seekers using online search, but this effect can be explained by the observable characteristics of the job seekers using traditional and online search. Kuhn and Mansour (2014) replicate this study with data from 2008 and 2009 and find that online job search significantly reduces unemployment duration. Bhuller et al. (2019) study how the roll-out of broadband internet affects labor market matching. Using the exogenous variation of the expansion of broadband coverage in Norway between 2000 and 2014, they find that broadband access is related to an increase in the number of firms posting jobs online. On average, the duration of a vacancy falls by 1% for a 10% increase in broadband coverage. Accordingly, the job finding rates of job seekers increase as well. Furthermore, job seekers with full broad band coverage enjoy 3-4% higher starting wages than those without, which supports the hypothesis, that better outside employment options increase the negotiation power of the job seeker. Similarly job tenures increase, which supports the idea that online job search improves the quality of job matches. Mang (2012) finds that the quality of job matches is higher when job seekers use online search. He studies employee data from Germany and uses the subjective evaluation of the employees as an indicator of match quality e.g. satisfaction, career perspectives, commute and ability to apply own skills etc. Kroft and Pope (2014) investigate the effects of online search by considering the entry of the website Craigslist to different regions in the US. They show that Craigslist increased the number of online search and reduced the amount of newspaper advertisements. While they can show that the vacancy duration of housing property has indeed decreased, they do not find any significant change in the unemployment rate. This is interpreted as an indication that search frictions in the labor market have not been lowered. Czernich (2011) studies the effect of broadband

access on unemployment rate in German municipalities between 2002 and 2006 and does not find a significant effect. One of the very few studies focusing on the effects of more sophisticated software is Horton (2017). He conducts a field experiment on a gig work platform. One group of employers receives algorithmically generated recommendations for job seekers to target, while the other group can only search for candidates themselves or wait for job seekers to apply. The study finds that especially employers in the technological field did follow the recommendations and had a 20% higher vacancy filling rate. The algorithmically recommended applicants had similar characteristics to the applicants which a firm would normally recruit. All of the papers considered so far, assumed that the use of different digital tools increases matching efficiency and found either positive effects or no effect. Another approach is carried out by Alexandrakis (2014) who analyzes the effect of IT use in production on matching efficiency with US time series data from 1967 to 2007. He focuses on the heterogeneity of IT use across firms when a new IT innovation is first adopted. At first, few firms adopt the new technology, but with time more firms adopt it as well. Therefore IT use heterogeneity is high in the early days of an innovation and declines with time. The innovations considered are mainframe computers in the early 1970s, PCs in the mid-1980s and the internet in the late 1990s. The author finds that increases in heterogeneity are related to declines in matching efficiency. The theoretical reasoning is that in the early adoption phase, firms start to require IT skills which most job seekers do not possess. The skill mismatch dissolves only with time, as the skill levels in the workforce adjust. Therefore, this author suggests that digitization could also decrease matching efficiency in the early phase of adoption, which makes research on the question even more interesting.

Overall, this literature review shows that other studies have also researched whether IT use in recruitment practices has increased matching efficiency. Existing studies have used a variety data types, methodologies and measurements of matching efficiency. However, the results are mixed and the number of studies is limited. Most studies have focused on internet job search, but very few have considered the effects of more recent innovations in recruiting. The majority of studies considers a short time span of only a few years. Therefore, my thesis aims to contribute to the existing literature by i) by focusing on Sweden, for which no such studies exist that I am aware of ii) by considering a longer time period (19 years), mostly past the year 2000 and iii) by focusing on overall digitization effects and not just on a specific innovation e.g. only online job search.

3 Data and Methodology

Identification Strategy Overview

I want to study whether digitization in recruiting has had an impact on matching efficiency. The next chapters will be structured according to three steps

1. Measure matching efficiency in the labor market regions (Chapter 4).
2. Measure firm digitization in the labor market regions (Chapter 5).
3. Study the relationship between changes in digitization and matching efficiency over time and ensure that the correlation can be interpreted as a causal effect (Chapter 6).

At first I will estimate matching efficiency in each labor market region, to obtain the independent variable for the regression in step 3. Second, I measure digitization, which will be my treatment variable in step 3. I assume that digitization in recruiting is closely related to the internal IT use in firms, which can be approximated as the number of employees with IT training. Furthermore, I assume that IT use in firms has not increased at the exact same speed across regions, but is driven by random variation, and by different pick up rates of IT processes between different industries. Since the industrial structure of the regions varies, this also leads to variation in IT adoption.[2] In the third step I estimate the relationship between digitization and matching efficiency. An identifying assumption for this step is that firm digitization is itself not influenced by matching efficiency. Furthermore, I assume that any positive correlation between the two variables can be attributed to the causal channel of recruiting. I focus on the changes within regions (using Fixed Effects and First Difference approaches) to avoid measuring whether region with overall higher digitization also have overall higher matching efficiency, since this is very likely to be confounded with other characteristics of the regions. I employ different robustness checks to avoid a bias caused by unobserved effects. Each chapter will elaborate the methodology and identifying assumptions for the respective step in more detail.

Data

The geographical units for my analysis are Funktionella Analysregioner (FA regions). These regions are divided by the Swedish Agency for Economic and Regional Growth (Tillväxtverket) based on commuting behavior between municipalities and represent

[2] This argument is inspired by the logic of a Bartik measure as outlined by for example Acemoglu and Restrepo (2020).

local labor markets (Tillväxtverket 2022). Therefore, FA regions are very well suited geographic units for analysing job matching efficiency. It is relevant to keep in mind that FA regions are of very different population size. Stockholm, Göteborg and Malmö far outweigh the other local labor markets. Therefore, the results of an unweighted panel analysis with FA regions should not be used to draw conclusions about matching efficiency in the aggregate Swedish labor market. The classification is updated every 10 years, based on the past and predicted changes of commuting behavior. In the time frame I am interested in, two classifications exist, 2005 and 2015. The classification from 2005 divides Sweden into 72 FA regions, while the new 2015 classification divides only into 60 regions. Thus, over time, people are more willing to commute longer distances. For a panel analysis it is necessary to use the same FA regions over the entire time period, otherwise I can not track the changes of regions over time. I decide to use the 2005 classification for two main reasons, firstly, the 2005 classification takes into account the predicted changes of commuting behavior until 2015 while also being timely closer to the earlier years of the sample. Secondly, having 72 instead of 60 regions allows to obtain more robust results because more variation and more clusters exist in the data. I will use the terms *region, local labor market and FA region* synonymous throughout this paper.[3]

For the first step, computing matching efficiency, I obtain my labor market data from the Swedish Public Employment Service (Arbetsförmedlingen). They provide monthly data of the stock of the unemployed, the newly unemployed, the newly listed vacancies, the stock of vacancies and the number of unemployed who found work during the month. The data is originally at the municipality (Kommun) level, which I aggregate to the level of FA regions. For the second step, measuring firm digitization, I require data on the number of individuals with IT skills. This data is obtained from Statstics Sweden microdata base LISA, which contains data on all individuals in Sweden from 16 years of age. It provides information on an individual's labor market status, municipality of living and working, education level and education subject and current job title. The data from this source contains yearly measures. I have access to this data for the time span 1998 to 2016 and use it to compute aggregated numbers for the FA regions. For the final analysis I also require data for covariates, which is obtained either from Arbetsförmedlingen for municipalities and aggregated to a FA level, or from the Micro data base LISA and also aggregated to FA level.

[3] Thus, whenever I use the term *region*, I do not refer to the Swedish administrative regions (län)

4 Job Matching Efficiency in Labor Markets

4.1 Methodology

As a first step, I need to find a suitable measure of matching efficiency. As I already touched upon in previous chapters, there are different ways of modelling and measuring matching efficiency. I need an approach which produces a measurement over time and across regions suitable for a panel analysis. In this section I will refer to the literature and successively develop my own approach.

4.1.1 Matching Efficiency Measurement in the Literature

A paper by Higashi (2020) is a very useful methodological orientation for a matching efficiency measurement. Higashi studies how the Tohoku earthquake in Japan has had differential effects of job matching efficiency between affected and unaffected regions. His research question is similar to mine, in that he is also studying the effects of an external shock which affected some Japanese regions more than others. Further, he is interested in the effect of this shock over time. In contrast to me, he uses a Difference-in-Difference setting. The dependent variable in his analysis is also matching efficiency. Higashi computes matching efficiency for each county at each point in time by taking the region and time specific residual of an estimated matching function. In the first step he estimates the matching function with regional panel data for Japan by employing a log linear form of the standard matching function.

$$lnM_{it} = \eta_1 lnU_{it-1} + \eta_2 V_{it-1} + \psi_i + \phi_t + \epsilon_{it} \tag{5}$$

He uses the stock of unemployment and vacancies from last period as an instrument for this periods stocks. ψ_i are region fixed effects and ϕ_t are time fixed effects. From this regression he obtains estimates for the parameters $\hat{\eta}_1$ and $\hat{\eta}_2$. Using these parameters, he then computes the residual of this matching function for each observation, that means for each region at each point in time. This residual, which he calls μ_{it} is a measurement of matching efficiency.

$$ln\hat{\mu}_{it} = lnM_{it} - \hat{\eta}_1 lnU_{it-1} - \hat{\eta}_2 lnV_{it-1} \tag{6}$$

Having obtained this residual, he then uses it as the dependent variable and carries out a classical Difference-In-Difference regression.

In my analysis I estimate the matching efficiency like Higashi (2020) with a slight alteration: I use a different functional form of the matching function, which will be explained in the next subsection.

4.1.2 Random versus Stock-Flow Matching

The simple matching function $M_{it} = U_{it-1}^{\eta_1} V_{it-1}^{\eta_2}$ upon which Higashi bases his analysis (in a log linear version), and which I discussed in Chapter 2, only considers the stocks of vacancies and unemployment. Modelling the function like this is in line with assuming the that the matching process between the vacancies and the unemployed is random, so-called **random matching**. The probability of finding a new job for an unemployed person is $\frac{m(U,V)}{U}$ and the probability of a vacancy to be filled is $\frac{m(U,V)}{V}$. Thus, there is no differentiation between new and old job seekers. The number of matches is assumed to consist of some job seekers who found a new job directly after becoming unemployed and of some job seekers who have been unemployed for a longer period of time. This is realistic to the degree that not all job seekers know of all vacancies immediately after becoming unemployed, but only learn about all vacancies over time. Therefore, some will randomly find work faster than others, irrespective of any personal characteristics (Petrongolo and Pissarides 2001).

In contrast, the theory of **stock-flow matching** assumes perfect information about vacancies by the job seekers. Newly unemployed individuals sample the entire stock of vacancies. Some find an acceptable job and immediately match, while the rest finds no acceptable job in the current stock of vacancies and remains in the unemployment pool. The group without a match enters the unemployment stock and samples only the newly incoming vacancies in the next period. Therefore, stock flow matching includes not only stocks, but also inflows of vacancies and unemployment into the matching function.

$$M = m(U, V, \acute{U}, \acute{V}) \tag{7}$$

V and U stands for vacancy and unemployment stocks, $\acute{U}$ and $\acute{V}$ represent the inflows of unemployed and vacancies. There is a multitude of empirical specifications of this function. The following is by Aranki and Löf (2008) who estimate time and region specific matching efficiencies.

$$M_{it} = A_{it} U_{it}^{\alpha_1} V_{it}^{\beta_1} \acute{U}_{it}^{\alpha_2} \acute{V}_{it}^{\beta_2} \tag{8}$$

Aranki and Löf (2008) also transform this function by taking logarithms to obtain a linear function.

Coles and Petrongolo (2008) compare random and stock flow matching and find more favorable evidence for stock flow matching. Forslund and Johansson (2007) study whether the Swedish labor market is better represented by stock flow or random matching and find much more favorable evidence for stock flow matching.

Based on this result, Aranki and Löf (2008) and Järvenson (2020) also choose stock flow matching as the functional form for estimating the matching function of the Swedish labor market.

4.1.3 Methodology for Measuring Matching Efficiency

Because the findings in the literature are more favorable of stock flow matching, I specify the matching function in this study according to stock-flow matching. Taking logarithms, I assume the following functional form of the matching function

$$ln\ M_{it} = \alpha + \beta_1\ ln\ U_{it}^{in} + \beta_2\ ln\ U_{it-1}^{stock} + \beta_3\ ln\ V_{it}^{in} + \beta_4\ ln\ V_{it-1}^{stock} + \mu_i + \delta_t + \epsilon_{it} \quad (9)$$

This functional regression form is the same as used by Aranki and Löf (2008). I estimate this function with data from the Public Employment Service. M_{it} refers to the number of job seekers who found work (matches) in region i in time t. This means that a match is only registered if the job seeker has been registered at the Public Employment Service. U_{it}^{in} refers to the newly registered unemployed (nyinskrivna arbetssökande) in the respective month, U_{it-1}^{stock} refers to the number of unemployed who are remaining in the unemployment pool at the end of the previous month (kvarstående arbetssökande). People between the ages of 16 and 64 can be registered as unemployed. V_{it}^{in} refers to the newly registered vacancies (nyanmälda vakanser) and V_{it-1}^{stock} refers to vacancies remaining at the end of the previous month (kvarstående vakanser). All of these variables are absolute numbers. For both, vacancies and unemployment, I use the stock from the period t-1, because these stocks represent the people and places which will be available for matching in the beginning of period t. α is a constant. I include region fixed effects μ_i to exclude biasing the matching function through region-specific (but time invariant) effects. Similarly I also include time-fixed effects δ_t to avoid any biases caused by time-specific effects (which are region-invariant). Moreover, I cluster standard errors at the level of regions to account for serial correlation.

I combine this functional form, with the methodology used by Higashi (2020). From the regression of this matching function I obtain the estimates for the coefficients $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and $\hat{\beta}_4$ and for the intercept $\hat{\alpha}$. Using these, I calculate the residuals of the matching function, which I denote with ω_{it}.

$$\omega_{it} = ln\ M_{it} - \hat{\alpha} + \hat{\beta}_1 ln\ U_{it-1} + \hat{\beta}_2 ln\ U_{it}^{in} + \hat{\beta}_3 ln\ V_{it-1} + \hat{\beta}_4 ln\ V_{it}^{in} \quad (10)$$

The residual represents the matching efficiency in region i in year t, since it measures the difference between the actual matches and the predicted number of matches, given the level of unemployment and vacancies. Therefore, the estimated matching

efficiency excludes all variation which is caused by cyclical changes as measured by unemployment and vacancy levels. ω_{it} is a measure of matching efficiency which still contains time and region specific variation.

$$\omega_{it} = \psi_i + \phi_t + \epsilon_{it} \tag{11}$$

4.2 Matching Efficiency - Results

4.2.1 Estimating Matching Functions

I first regress the matching function with the monthly labor market data of 72 FA regions.

$$ln\ M_{it} = \alpha + \beta_1\ ln\ U_{it}^{in} + \beta_2\ ln\ U_{it-1}^{stock} + \beta_3\ ln\ V_{it}^{in} + \beta_4\ ln\ V_{it-1}^{stock} + \mu_i + \delta_t + \epsilon_{it} \tag{12}$$

For the unemployment measures U_{it}^{in} and U_{it}^{stock} I vary between using the total number of registered job seekers and the openly unemployed. Total job seekers include all people who are registered as job seekers by Arbetsförmedlingen. The openly unemployed are a sub group of total job seekers. These are job seekers who are not enrolled in any labor market program, they are actively looking for work and would be able to start at a new job immediately. Over the time period 1998 - 2016, the openly unemployed comprise on average 32% (min. 24% and max. 41%) of the total job seekers (Arbetsförmedlingen 2022). Additionally to the number of open unemployed, the group of total job seekers also includes people who are enrolled in labor market programs run by Arbetsförmedlingen, e.g. internships, trainings preparing for job entry; people who have work and do not obtain government benefits, e.g. part time unemployed and people whose employment will end soon; people who work and obtain government benefits e.g. people working in government subsidized jobs; and other job seekers. Therefore the total job seekers are including a very large group of job seekers, of which many can be assumed to be not as actively looking for a job as openly unemployed because they are in some form of program (Arbetsförmedlingen 2022). Generally, both open unemployed and total job seekers could be used to estimate matching efficiency. Total job seekers comprise the highest possible number of job seekers, which probably overestimates the number of persons actively looking for a job. In turn matching efficiency would be underestimated, because the number of matches is compared to a larger number of job seekers than there actually is. Openly unemployed are in contrast a very narrow definition, underestimating the number of actual job seekers and therefore overestimate matching efficiency. It should also be noted that the matches M are all matches of all registered job seekers who found work during the month. Using only the open unemployed in the regression therefore relies on the assumption that most of the registered matches can be

attributed to openly unemployed job seekers. Therefore, total and open unemployed can be seen as the maximum and minimum approximation of the actual amount job seekers, resulting in maximum and minimum approximations of matching efficiency. In section 6.1.2, I will further elaborate the advantages and disadvantages of using either of these two measures.

Table 1: Regression of Matching Functions 1998 - 2016

	Total Unemployed	Total Unemployed	Openly unemployed	Openly unemployed
$ln\ U_{it}^{in}$	0.2600***	-0.0148	0.1332***	-0.0638***
	(0.0229)	(0.0224)	(0.0205)	(0.0206)
$ln\ U_{it-1}^{stock}$	0.6219***	0.7561***	0.2021***	0.3531***
	(0.0437)	(0.0369)	(0.0369)	(0.0295)
$ln\ V_{it}^{in}$	0.0996***	0.0471***	0.1099***	0.0550***
	(0.0085)	(0.0068)	(0.0096)	(0.0075)
$ln\ V_{it-1}^{stock}$	0.0533***	0.0163***	0.0524***	0.0223***
	(0.0089)	(0.0048)	(0.0082)	(0.0052)
Constant	-1.6308***	-0.8040**	2.7443***	3.0991***
	(0.3077)	(0.3092)	(0.2419)	(0.1919)
Month Fixed Effects	-	Yes	-	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.9581	0.9790	0.9535	0.9772
Number of observations	16416	16416	16416	16416

Note: Standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

The results of the regression can be found in Table 1, column 1 and 2 represent results with unemployment measured by the total number of registered job seekers, in column 3 and 4 the number of openly unemployed is used. Assessing the effects for the regression with the total unemployed without season controls, I find that all explanatory variables have a positive effect on matches. The stock of unemployed has by far the largest impact and inflows of vacancies are slightly more relevant than the stocks of vacancies. In each regression I control for region and year fixed effects. When additionally controlling for season fixed effects (months), the coefficient of the unemployment inflow turns insignificant and slightly negative, but the other results are broadly similar. Compared to that, regressing on the openly unemployed, I still find that the stock of unemployed has the largest impact, but this impact is comparatively lower and the other variables have gained importance. When accounting for season effects, the inflow of unemployed has a negative and statistically significant coefficient. However, this coefficient is not very robust and turns statistically insignificant when using a slightly longer time period from 1996 - 2021 for the regression (see Table A12 in Appendix 1). For all regressions, the R^2

is above 0.95, so the estimated model is explaining most of the variation observed in the data, which can be expected given the use of fixed effects.

Comparing my results with other estimates of Stock-Flow matching in the Swedish labor market, I find some similarities and some differences. Aranki and Löf (2008) regress the matching function with data of the 21 Swedish regions (län), with monthly data from The Swedish Employment Agency over a time period from January 1992 until September 2007. Including region, year and season fixed effects, they also find that the impact of the unemployment stock has the highest impact on matching. Moreover, they also find that the inflow of vacancies has a larger effect on matching than the stock of vacancies. In contrast to my result, the inflow of unemployed has a relatively small positive coefficient and is still statistically significant at the 5% level. All coefficients have a positive impact on matching in their regression. Thus, my results are broadly in line with theirs, except for the unemployment inflow variable. Järvenson (2020) uses a very similar approach, with monthly data from Swedish regions (län) over the time period 2000 until 2018. When including region, year and season dummies, he also finds that the stock of unemployed has the largest impact on matching. Differing from Aranki and Löf (2008), the inflow of unemployed has however a larger impact than the inflow of vacancies. Järvenson (2020) in contrast to my results and those of Aranki and Löf (2008) finds a statistically insignificant, very small and negative effect of the stock of vacancies.
So my results are in line with the two authors in the main points: the stock of unemployed is the most relevant variable, the inflow of vacancies are more relevant than the stock of vacancies and the coefficients are of similar magnitude.

4.2.2 Matching Efficiency as the Residual

Using the regression results, I compute the time and regions specific residuals of the estimated matching function ω_{it}.

$$\omega_{it} = ln\ M_{it} - (\hat{\alpha} + \hat{\beta}_1 ln\ U_{it}^{in} + \hat{\beta}_2 ln\ U_{it}^{stock} + \hat{\beta}_3 ln\ V_{it}^{in} + \hat{\beta}_4 ln\ V_{it}^{stock}) \tag{13}$$

This ω_{it} represents time and region specific matching efficiency, based on the monthly labor market data and the estimated coefficients of these variables from the matching function. Since the digitization measures are only observed on a yearly basis, I compute a yearly matching efficiency by taking the average of the 12 monthly matching efficiencies for each year for each region. The matching functions with seasonal controls yield higher variation in the residual than the matching functions without seasonal and only year controls (appendix table A13). Given that seasonal controls add a further explanatory variable to the regression, a lower variance of the

residuals could be expected. Therefore, this behavior indicates that seasonal controls are very relevant to include, which is intuitive as labor markets typically experience seasonal fluctuations.´Table 2 shows the summary statistics of the yearly averaged residuals for open and total unemployed when seasonal controls are included. As expected, the mean of the residuals is practically equal to zero. Using the openly unemployed numbers yields a higher variation among the residuals than using the total job seekers numbers. But this difference is mainly driven by a higher variation between regions. Matching efficiency variation within regions is overall smaller and more similar between total and open unemployed.

Table 2: Within and Between Variation of Matching Efficiency

	Variation	Mean	Std. dev.	Min	Max	Observations
Open Unemployed	overall	0.000	0.8009249	-2.372055	2.335396	N = 1368
	between		0.7751004	-1.502945	2.086345	n = 72
	within		0.2204786	-1.047845	0.6991239	T = 19
Total Jobseekers	overall	0.000	0.2718325	-1.094925	0.9018487	N = 1368
	between		0.2125686	-0.5014374	0.6516204	n = 72
	within		0.1711796	-0.8326239	0.4076131	T = 19

Note: Residuals from the regression with seasonal controls.

Figure 1 shows the scatter plots of the yearly matching efficiencies as obtained by the different regressions. We can observe that matching efficiency has been following a slightly negative trend over time. This result is in line with the findings of other studies on the Swedish labor market. It can be observed that there is a slight dip around 2008, this is also in line with the literature (Eklund et al. 2015). The overall pattern could be related to institutional reforms in Sweden and a higher portion of people sent to labor market programs as I explained earlier.

Overall, my findings on the development of matching efficiency are in line with the literature. The estimated measure of matching efficiency will be used as dependent variable in the panel regressions in chapter 6. When using this measurement of matching efficiency it is relevant to keep in mind that this measure is focused on the quantity of matches. This is inspired by search theory as I outlined in chapter 2, but does not include information of the quality of matches. It does not measure the fit between the job and the hired candidate, e.g. measured by the length of the subsequent employment. While digital tools in recruiting might also affect the quality of the matches, this aspect will not be assessed in this study.

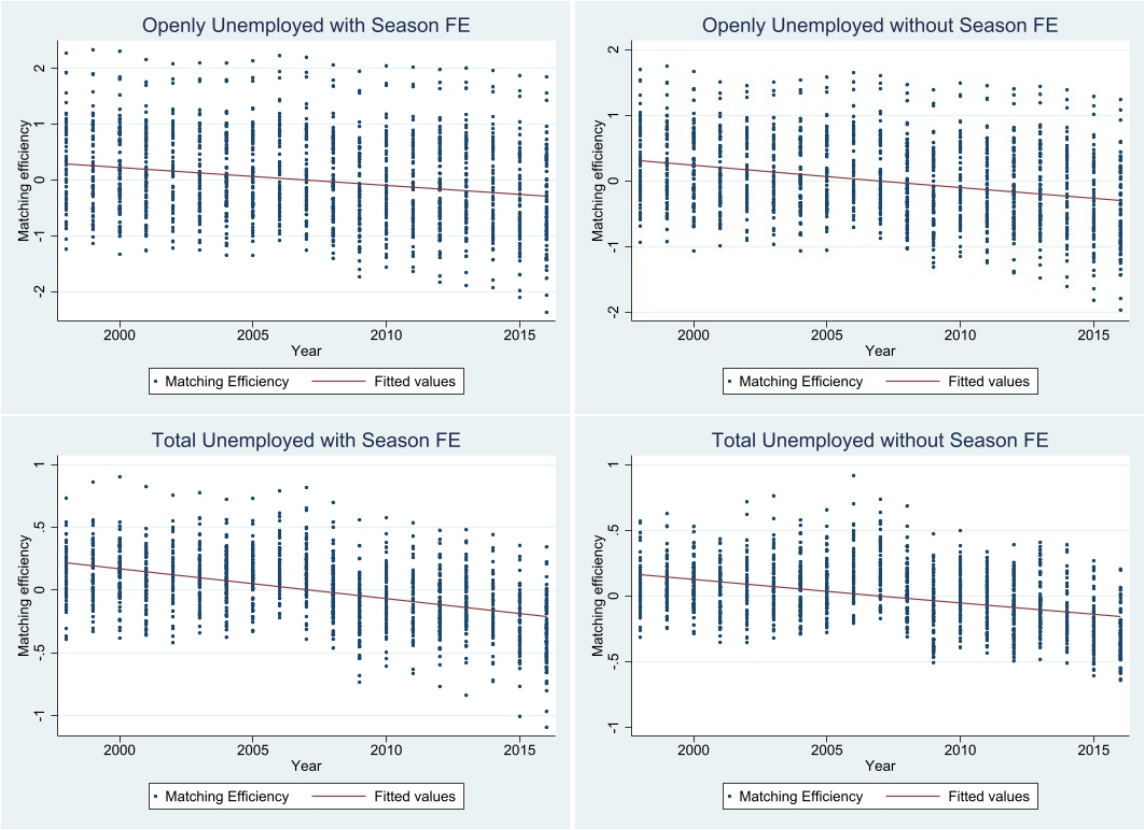


Figure 1: Matching Efficiency, Residuals of Different Matching Functions

5 Firm Digitization in Labor Markets

In this chapter I describe how I measure digitization in firms, name the methodological limitations of this measure and give an overview over the results of this measurement. The digitization measure constructed in this chapter will be the main explanatory variable in my panel regression in chapter 6.

5.1 Methodology

Since the main causal mechanism which I consider is IT use in recruiting, I would ideally measure how firms use digital tools in recruiting. Unfortunately, I do not have access to such data. Therefore I have to use an approximation variable. I assume that in order to use digital technologies it is necessary to have personnel which is familiar with information technology and able to implement new technologies. Further, I assume that firms using more digital technology overall, will also use more digital technology internally, for example in recruiting. Therefore, I consider two possible proxy variables, the total number of working individuals in each region who either work in an IT related job or who have an IT related education. I will analyze and compare these two numbers empirically in the second section of this chapter (5.2). In the remaining part of this methodology section I discuss the conditions of using either one of these measures as a proxy variable.

The unobserved variable is the actual internal IT use in firms and in recruiting specifically. I replace this variable with the approximation variable. In any possible regression I will therefore put in proxy x* instead of unobserved x. Thus, a hypothetical regression would be $y = \beta_0 + \beta_1 x^* + \beta_2 z + e$. For this to be econometrically valid the following conditions must be fulfilled (Wooldridge 2018, p.300):

1. The proxy and the unobserved variable must be correlated $x^* = \alpha_0 + \alpha_1 x + v$
2. The error term e needs to be uncorrelated with both x* and x, the proxy and the unobserved variable.
3. The error term v needs to uncorrelated with z, in other words $E(x^*|x, z) = E(x^*|x) = \alpha_0 + \alpha_1 x$

Condition 3) entails the assumption that the correlation between the use of IT in recruiting and the number of IT specialists does not depend on any other control variable which I include in my regressions in chapter 6. Due to my complete lack of data on the IT use in recruiting I can not prove or disprove whether this assumption

holds, it is therefore a relevant limitation of my method to keep in mind. Condition 2) is very similar to the general OLS condition and thus relatively uncontroversial. Condition 1), the relationship between proxy and unobserved variable will be discussed in more detail in the remainder of this section.

How reasonable is it to assume that the number of IT specialists is correlated with the internal IT use in firms? Generally, I can only argue but not show that this proxy is reasonably correlated with IT use in the firm, because I do not have access to the relevant data. On the one hand firms which use more IT internally should also employ more IT experts. Firms can obtain IT products and services by producing them internally, for which they need personnel, but they can also buy such services externally. Considering data from SCB in 2016, I observe that a higher number of firms obtains IT products and services externally than produces them internally. Across all IT categories, around 50% of firms obtain services and products externally. Support of public office software is internally produced in 43% of firms, development of business systems is done internally at 32 % of firms, support for business systems in 21%, web solution development in 27%, web solution support in 20% [4] However, when firms are sorted by firm size, it can be observed that the numbers increase steeply. For firms with more than 50 employees, 21 % of all firms provide provide support for business systems internally but 29% of firms with 50 - 249 employees, and 39% of firms with over 250 employees. In firms with more than 250 employees the internal production of IT services and products is rather high, 61% provide support for public office software internally, 57% develop business systems and 46% develop web solutions (Statistics Sweden 2022b). This perspective is very relevant since larger firms account for a much higher amount of employees and thus for more vacancies than small firms. In 2003, 27% of employees in Sweden worked in firms with less than 50 employees, while 61% worked in firms with more than 200 employees and 53% worked in firms with more than 500 employees. In 2013, the number of people working for small firms had increased to 36%, but still 64% work for firms with more than 200 employees (Statistics Sweden 2022a). Another survey asked firms whether they obtained AI by own development, own modification of external software or fully by external providers. Across Sweden, 8% of firms obtained it by own development or own modification, while 9% obtained it only via external providers. Overall, we can observe that internal production of IT products and services plays a significant role (Statistics Sweden 2022d).

Summing up, I argued that IT products and services are internally produced at

[4] These numbers do not add up, indicating, that some firms might not employ some of these at all, especially for the support for business systems or for web solutions, 30% of firms did not indicate any provider.

a high share of firms, especially in larger firms which have more influence on the recruitment practices in the overall labor market because they account for a larger share of the vacancies. When IT is internally produced, it can be assumed that the firms employ skilled personnel for these tasks. Additionally, even when a service or a product is bought externally, there is a high chance that there will still be an employee in the firm who is responsible for administering these products and has some IT knowledge. Therefore, while it might not hold in every case, it is reasonable to assume, that firms who are using more IT will on average be also more likely to employ more IT specialists.

On the other hand, the correlation should work in the other direction too. Do firms with many IT specialists also use a lot of IT products internally? Taking this point of view, firms also employ IT specialists for non internal purposes, such as developing products for their clients or for operating IT services for other firms. While I acknowledge that a certain share of the IT specialists in firms was probably not hired for internal IT production, I would argue that the presence of these IT experts in the firm is likely to create spillover effects to other departments and employees. Employees with a high awareness and knowledge on digital tools will probably not only develop IT, but also use more sophisticated digital tools in their own internal work flow and push the firm to obtain more software etc. Awareness and knowledge about these work practices is more likely to spread among other non-IT employees if some digital pioneers are working in the firm. Awareness and understanding of digital tools and technologies is one of the most relevant barriers for using them, additional to cost concerns or the compatibility with current processes and technologies. This is confirmed by a more recent statistic by Statistics Sweden, which asks firms for the reasons they refrain from using AI. Across all Swedish companies, the lack of relevant expertise at the enterprise is the most named reason, being named twice as often as cost concerns (Statistics Sweden 2022c).

Overall, I have argued that it seems plausible to assume that firms with more IT specialists use more IT in their internal processes and that firms which use more IT internally employ more IT specialists. The limitation remains that I can not prove this relationship and moreover can not measure how closely these two variables are correlated. Furthermore, I assume that if IT is used in internal processes in general it is also used in recruiting, but I can not prove this either.

5.2 Firm Digitization - Results

In this section I focus on measuring the number of IT specialists who work in each region in each year. The data is obtained from Statistics Sweden micro data base

LISA. As mentioned earlier, there are two options for classifying IT specialists, either by a formal education in an IT related subject, or by an IT related job title. For both measures, I count the number of individuals by the municipality of their workplace and I only count individuals who work, irrespective of the form of work. Thus, I exclude people who are unemployed or retired, because they do not use their IT skills in the economy. Furthermore, persons without work do not have a job title, which becomes problematic when comparing the two measures, education and job title. In the remaining part of this section I will explain how I measure IT specialists according to each classification. Afterwards I compare the numbers obtained by the two measures.

At first I focus on the classification of an IT education. Statistics Sweden uses a classification for education subjects called Sun2000Inr. In this classification each education subject is given a code, all codes starting with 48 comprise educations in the field "Data". These are more narrowly broken down into "General education" ("Data, allmän utbildning"), "Data Science and Systems" ("Datavetenskap och systemvetenskap"), "Computer application" ("Datoranvändning") and "Other". I define an IT education as an education in any of these subjects. While the subject classification does not entail any information about education levels itself, information about education levels exists in the data set. I observe that persons with an IT education have education levels ranging from secondary education up to research levels (level 3 to 6 in SCBs Sun2000niva classification). Thus, IT education include for example vocational training, a university degree, a PhD or persons who participated in any of those education programs but who did not finish with a degree.

The number of people with such an education in an IT subject has increased sharply over the years. In 1998 there were 38 258 people with an IT education working in Sweden. This number has increased gradually over the years and stood at 93 034 in 2016. Considering the different education levels, all levels of IT education have increased in absolute numbers: secondary education, post secondary (less than two years), post secondary (more than two years) and research education. The shares of different education levels among the total amount of IT educations stayed roughly similar over time (1998-2016). In 2016, 58% of all working people with IT education had participated in post secondary training which was longer than two years, around 20% participated in post secondary education shorter than two years and 20% in secondary education (vocational training), 2% undertook research education.

The second option to measure the IT related skills in the workforce, is by counting the number of people who have an IT related job title. Collection of job title data starts from 2001 with the classification SSYK4, from the year 2014 onward, the

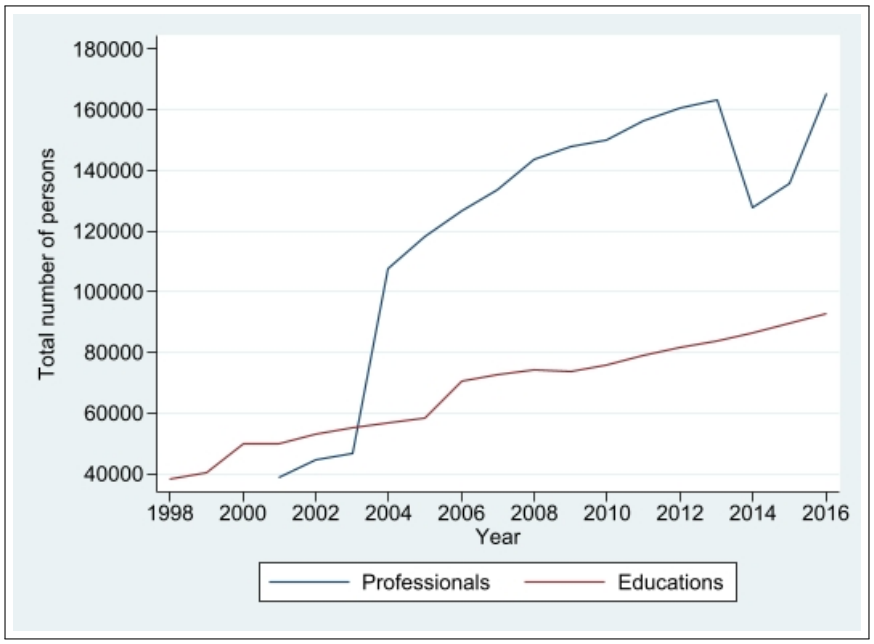


Figure 2: Number of IT specialists, Total, Sweden

updated classification SSYK4_2012 is used. In this classification each job title has a code. The exact codes which I used to define IT jobs can be found in Appendix 2. The job titles include information about responsibility and education requirements as well. All considered IT job titles indicate a requirement for some form of higher education.

The number of people with an IT related job title also grew sharply, in 2001 38,909 persons in Sweden worked in an IT job, in 2010 it were 150,137, in 2016 it were 165 426. An anomaly in this rise, is the introduction of the new classification in 2014, in this year I observe a sharp drop in the number of IT professionals, from 163 295 in 2013 to 127 551 in 2014. The numbers continue to rise in the subsequent years. This drop seems to be related to the change in classification, although the job codes which I used for the old and new classification should be equivalent according to Statistics Sweden. A problem of the measurement of job titles, is that there are considerable amounts of data missing for this measure. Confining the focus only to the working population, in 2001, 49.26 % (2.4 Million) of the persons in the statistic have an unknown job title. This drops to 12.65% (670 000) in 2010 and rises again to 15 % in 2016 (790 000). In contrast, for education the share of missing information is relatively low and steadier, 2.6 % of persons have an unknown education in 2001, 2.9% in 2010 and 3.0% in 2016.

Figure 2 compares the development of the total number of persons with an IT education (Educations) and with an IT job title (Professionals). The development of

the education numbers reflects a steady growth. The number of IT professionals is lower than the number of educations in the beginning but then increases sharply and overtakes the number of educations by far. Given that a large part of the job titles are missing, the number of people working in IT must far outweigh the number of people having an education in IT. However, it can also be observed that the number of IT professionals is behaving much more volatile, which is probably driven by the changing amount of missing data and the change in classification.

Given the more steady development, it seems preferable to use the number of people with IT education rather than the number of IT professional. Education seems to underestimate the actual number of people with IT skills in firms, but this is less problematic if it is assumed that education and the actual number of IT professionals (without measurement error) show a constant correlation.[5] While I can not check this assumption, since I do not observe the actual number of IT professionals, I can at least estimate the correlation between the measured IT educated persons and IT professionals within each region in each year.[6] I do find a highly significant positive correlation coefficient (Table A14 in Appendix 1). This indicates that there is a robust correlation between the two measures within years and regions and that the number of professionals is on average 2.8 times as high as the number of IT educated. The R^2 is around 0.66, indicating that most of the variation in the number of professionals can be explained by the number of IT educated persons. To observe the behavior of the correlation between the two measures over time, I also run the regression for each year separately and compare how the coefficients differ. As expected, the coefficients do vary over the years but they are all statistically significant, which indicates that the two measures are correlated in each year (Appendix Tables A15, A16, A17).

The number of IT educated individuals in region i in year t will be used as the explanatory variable of interest in the panel regression in the next section.

[5] Since the measurement of professionals in the data contains measurement error, the correlation between education and professionals in my data can not expected to be constant.

[6] I include year and region fixed effects to compare whether these two measures are correlated within the municipalities for each year separately.

6 The Effect of Digitization on Matching Efficiency

6.1 Methodology

After having obtained a measurement of matching efficiency and of digitization, the third step is to analyze how these two variables are correlated and to infer a causal relationship as far as this is possible. In the methodology part of this chapter I will first describe the functional form of my analysis and its strengths and limitations, second, I will point out factors which could confound this analysis and explain which controls I propose accordingly.

6.1.1 Functional Form of the Regression

The most basic regression equation to estimate the correlation would be

$$\omega_{it} = \alpha + \gamma_1 ln\ Digi_{it} + \gamma_2 ln\ C_{it} + \epsilon_{it} \tag{14}$$

ω_{it} is the computed matching efficiency for region i in year t, Digi stands for firm digitization and C is a vector of control variables, which I will specify in the second part of this section (6.1.2). I do not take the logarithm of ω, because ω is obtained as the residual of a logarithmic regression and because ω also takes on values below zero. However, this very simple regression is unlikely to yield the estimates I am interested in. A simple OLS regression mixes two effects, how an increase in digital skills within a labor market region will impact matching efficiency within the region, and how different levels of digital skills explain differences in matching efficiency between labor market regions. I am not interested in the later, whether regions with a higher level of IT specialists have overall better matching efficiency, since this effect is presumably confounded in many ways. The level of digital skills can be expected to vary substantially between regions due to different industry structures. A region with a high share of IT industry will have a generally higher level of IT specialists. Regions with a high share of IT industry might be systematically different from other regions in observable and unobservable characteristics. For the OLS estimator to be unbiased, the explanatory variables have to be uncorrelated with the error term (Wooldridge 2018, p. 460), but this condition would be violated, if digital skills are correlated with the unobserved region effects in the error term. Therefore, I will focus only on the effect of digitization on matching efficiency within regions. Suitable approaches for this are fixed effects and a first difference design, which I will evaluate separately in the following paragraphs.

Fixed Effects (FE)

When employing region fixed effects, regressions are run on the deviation from region averages of variables instead of the variables themselves. Thus unobserved region specific effects, denoted by μ_i are eliminated and the estimated effect will rather reflect the within variation in each labor market region.[7]

$$\omega_{it} - \overline{\omega_i} = \alpha - \overline{\alpha} + \gamma_1(ln\ Digi_{it} - \overline{ln\ Digi_i}) + \gamma_2(ln\ C_{it} - \overline{ln\ C_i}) + \mu_i - \overline{\mu_i} + \epsilon_{it} - \overline{\epsilon_i} \quad (15)$$

Another way to obtain results equivalent to the fixed effects regression is by including a dummy variable μ_i for each region i. This yields the same coefficients and standard errors as a time demeaned regression (Wooldridge 2018, p.488).

First Difference (FD)

Another option for focusing on the changes within each region is the first difference approach, which also eliminates region specific unobserved characteristics. The basis for taking differences is the previous fixed effects equation

$$\Delta\omega_{it} = \Delta\alpha + \Delta\gamma_1 ln\ Digi_{it} + \Delta\gamma_2 lnC_{it} + \Delta\gamma_3\mu_i + \Delta\epsilon_{it} \quad (16)$$

Assuming Δ captures a one period change between year t and year t-1, we can eliminate some variables from this equation. α and μ_i will drop out, since they are constant over time.

Disadvantages of FE and FD

In this paragraph I consider general disadvantages of these two methods and consider their relevance for my research question. One disadvantage is that for all variables with relatively small changes over time, much of the information will be eliminated. I am mainly interested in the effect of the number of individuals with IT education, this variable does however exhibit some significant change over time, as I observed in the chapter 5. Therefore, I am less concerned about this limitation.

Another related issue is a high sensitivity to measurement error. There is a risk that some part of a variables' variation over time is caused by measurement error. If measurement errors account for a large share of the variation over time, the estimates can be biased, since only the variation over time remains in the regression (Angrist and Pischke 2008). Measurement error can not be excluded for the digitization variable and other control variables. However, the data I use is based on registry data, which typically entails far less measurement error than for example

[7] Please note that, while I still use the same letters for the coefficients α, γ these coefficients are different from the the coefficients of the simple OLS regression.

survey data. Also, my dependent variable ω is computed as the residual of the matching function, which takes in the number of matches, unemployed and vacancies as registered by Arbetsförmedlingen. The variables used for this computation are likely to contain some measurement error, since not every job seeker and not every vacancy gets registered there. While this is a little bit problematic, I am not too concerned about this drawback, since I can expect this measurement error to affect the number of matches, unemployment and vacancies evenly. Furthermore, all data points in all my regressions are based on aggregated micro data. Individual measurement errors on the micro level are likely to exist, but also to cancel each other out when the data is aggregated. Thus, as long as these measurement errors are random, it should not bias the results.

A general problem in panel data regressions, is that error terms are very likely to be serially correlated. This follows from the nature of time demeaning (FE) or differencing (FD). The error term of the FD regression is the difference of the population error. If the differenced error term is serially correlated, the population error term has to be serially uncorrelated (Wooldridge 2018, p. 470, 490). The same holds for FE, the error term of the FE regression is the time demeaned population error. If the population errors are serially uncorrelated, then the time demeaned errors have to be serially correlated (Wooldridge 2010, p.270). Generally, serial correlation is a violation of basic OLS assumptions which state that individual observations need to be independent (Wooldridge 2018). While there might be no ideal way of dealing with this, Angrist and Pischke (2008) explain that adjusting for geographically clustered standard errors is a sufficient way to account for this. To be valid, this approach requires a sufficiently high number of clusters, which the authors determine to be at least 42 clusters. I have 72 regions in my sample to cluster on. Wooldridge 2018, p.483 also states that clustering standard errors on the individual cross section unit is a sufficient way to deal with serial correlation and heteroskedasticity in the error terms of panel regressions. Therefore, I generally cluster standard errors at the region level in all regressions.

Fixed effects and first difference regressions do not yield the same results for more than two time periods. Both approaches are unbiased and consistent. Which one is more efficient dependents on the behavior of the error term. If the population error terms are serially uncorrelated, fixed effects are more efficient than first difference. We can not directly observe the behavior of the population error term in the FE or FD regression, but we can infer that the population error is serially uncorrelated if the differenced error term is serially correlated. Serial correlation can be detected by testing a simple auto regressive process, for example $\Delta\epsilon_{it} = \Delta\epsilon_{it-1} + s_{it}$ (Wooldridge 2018, p. 470, 490).

6.1.2 Confounding Factors and Included Controls

In this section I discuss limitations of the empirical strategy and focus on different factors which influence matching efficiency apart from digitization. I explain how I try to account for these aspects in order to avoid biases. I am only concerned with confounding effects related to the changes of matching efficiency within the same region since I always control for unobserved region effects. In the first two paragraphs I will discuss concerns related to my main explanatory variable of digital skills, in the later paragraphs I elaborate on factors which influence of matching efficiency in general and explain how I control for them.

Simultaneity bias

A general concern in regression analysis is simultaneity bias or reverse causality, which means that the explanatory variable is impacted by the explained variable. In this regression it would mean that the number of IT specialists is itself influenced by changes in matching efficiency. A possible causal mechanism could be that firms are more able to recruit IT specialists in regions in which matching efficiency has relatively improved. One possible explanation of the decline in matching efficiency in Sweden, is that small local labor markets have a harder time to recruit specialists from other regions, because they do not offer enough employment opportunities for an accompanying spouse (Eklund et al. 2015, p. 22). If that was the case, better job matching efficiency could result in better opportunities to recruit IT specialists. In this case we would face a reversed causality. While the possibility of this mechanism is a limitation for the findings of any regression between digital skills level and matching efficiency, I do not think that it is a strong limitation. The outlined argument is purely hypothetical and is not widely discussed by other authors. Besides that, the argument only applies to a certain subgroup of IT specialists, namely those who live in an urban region and consider moving to a rural region and who have a spouse working in a quite specialized profession to accompany them. Therefore, I presume the risk of simultaneity bias to be low. To minimize the risk of reversed causality econometrically, I perform regressions with lagged explanatory variables as a robustness check of my results in section 6.2.4.

Causal channels unrelated to recruiting

It is important to keep in mind that the digitization variable only approximates the internal IT use and actually measures the number of IT specialists. I assume that more IT specialists lead to more internal IT use and to more IT use in recruiting. But more IT specialists could also impact matching efficiency via other channels.

Another possible causal channel, which I mentioned in the very beginning, is that greater IT use leads firms to require a higher level of IT skills of new applicants for positions on all job levels, e.g. for low, medium and high skilled jobs. Thus, there could be greater skill mismatch between the skill requirements of vacancies and the skills possessed by job seekers after IT innovations were adopted (Alexandrakis 2014). The argument implies the assumption that a large share of vacancies require IT skills, otherwise the skill mismatch would only affect a small portion of the vacancies and be unlikely to have a significant effect on matching efficiency. Since I do not possess any data on the amount of vacancies which require IT skills for the respective time in Sweden, I can not verify how likely this argument is to apply. This hypothesis would predict a decline in matching efficiency, contrary to the prediction I propose. If both causal channels apply to a certain degree, the regression results will be determined by a mix of these two effects. Since the two effects are contrary, the result would still be an indication which effect is more relevant.

Another causal channel unrelated to recruiting could be productivity effects. Productivity changes could influence matching efficiency but can also be directly linked to digitization in firms. There are two possible relations between matching efficiency, digitization and productivity.

1. The increased use of ICT in firms increases production productivity and higher productivity could have effects on matching efficiency

2. ICT use in recruiting improves matching efficiency and better matching efficiency improves productivity in production

Both causal channels would predict a positive correlation between digitization and matching efficiency. The first channel would however be unrelated to recruiting, which is the causal channel I have assumed. Therefore, in the first case, controlling for productivity growth would be essential when the goal is to focus purely on the effect of digital recruiting. But in the second case, productivity controls would at least partly control away the effect of digitization and thereby downward bias the correlation coefficient. Therefore, not adding controls for productivity, does not risk a bias in the results, the only risk is, that the causal channel of the correlation might not be recruiting but productivity. I consider which mechanism seems more likely: Whether ICT use improves production productivity is not unanimous in the literature, but I assume for now that this relationship exists.[8] The first argument requires further that productivity also impacts matching efficiency. Ladu (2012) find

[8] It has been observed by some studies, that increased ICT use is associated with higher TFP. The divergence of productivity between the US and Europe can be partly explained by this (Biagi 2013). However, for example Shea (1998) finds that technology shocks do not increase TFP.

that the relationship between productivity (measured as TFP) and labor market variables, e.g. employment, seems to be ambiguous. Overall, there is much more literature focusing on the reverse relationship, the effects of matching efficiency (turnover, employment and unemployment etc.) on TFP. Many of those studies did find significant effects (Ilmakunnas et al. 2005, Mukoyama 2014). Therefore, while both arguments seem possible, I find the second argument slightly more plausible. Thereby, the risk that an estimated correlation is mainly driven by productivity and not by recruiting does not seem very high.

However, for other reasons, which will be elaborated in the next sections, I also include a region-specific and a general time trend in the regression, which would also control for unobserved gradual (long term) productivity changes. When controls for year fixed effects are added, they might control unobserved productivity changes even more granular than a time trend would do. Therefore, it is relevant to keep in mind that these controls, especially the year fixed effects, might downward bias the estimation result if argument 2 applies, namely when production productivity increases after digitization increased matching efficiency.

After having elaborated the limitations which are directly related to the digitization variable, the remainder of the section will focus on more general factors which influence matching efficiency and are therefore relevant to control for to exclude a spurious regression.

Dispersion of employment and structural shift

Many studies consider changes in the location and dispersion of employment opportunities to be potential causes of the changes in aggregate matching efficiency. The most prominent argument is that of sectoral shift: Different industries exhibit different employment developments, some may increase or decrease their labor demand. Since industries are clustered in different geographies, sectoral shift changes the geographic allocation of labor demand and the employment shares of industries in each region. This has an effect on matching efficiency, since job seekers might be underrepresented in regions with increasing labor demand, since they live in other regions and face mobility constraints. Furthermore, sectoral shift changes the relative importance of industries in the labor market within each region. Assuming that different sectors exhibit different matching efficiencies, this affects overall matching efficiency in the region. Furthermore, different industries have different skill requirements, the skill requirements for new jobs in a region change when the industrial structure in the region changes. Thus, sectoral shift could cause a skill mismatch and lead to deteriorating matching efficiency (Bonthuis et al. 2016, Lazear and Spletzer 2012, Petrongolo and Pissarides 2001, Barnichon and Figura 2011).

Therefore, I account for structural shift with a region-specific time trend $\mu_i * t$.[9] Including this, every region gets an own time trend coefficient in the regression. These individual time trends account for general, long term changes in matching efficiency in that region over time. Since structural changes are presumably long term rather than short term processes, this seems to be a sufficient control.

Cyclical shocks

Cyclical shocks should in theory not influence matching efficiency, since they only cause shifts along the Beveridge curve but not of the curve itself. The same holds for matching efficiency estimated by a matching function, since unemployment and vacancy variables control for such cyclical shocks. Nevertheless, some studies do find significant effects of cyclical shocks on matching efficiency. Bouvet (2012) finds that Total Factor Productivity (TFP) growth and a positive output gap are associated with inward shifts in the Beveridge curve. Similar, Börsch-Supan (1991) finds that shifts in the Beveridge curve can not fully be explained by structural parameters such as composition of the unemployment pool. In his research cyclical parameters, can indeed jointly explain a proportion of the shifts. He uses for example GDP per capita as an additional control.[10]

Despite these findings there is no unanimous consent that cyclical shocks (additional to the cyclicality captured by unemployment and vacancies) do actually affect matching efficiency. Bonthuis et al. (2016) find that cyclical effects themselves are of minor importance for explaining shifts in the Beveridge curve. Other studies do not even include cyclical factors in their estimation, especially those which are, like me, interested in regional differences of matching efficiency in the same country, e.g. Coles and Smith (1996) and Pedraza (2008). Furthermore, the mentioned authors who find effects of cyclical measures, Bouvet (2012) and Börsch-Supan (1991), estimate Beveridge Curves which only take into the stocks of the vacancies and the unemployed. These are already strongly cyclical measures, however, I use a

[9] While it might be preferable to have a direct measurement of structural shift for each Swedish labor market region, this is hard to obtain due to limitations in data availability. The main issue is that the classification for the industry structure SNI has been modified dramatically in 2007. With this it becomes very difficult to assort the equivalent classifications before 2007 and after 2007. To include the structural shift measure in the regression it would be important to have the same industry measures for the entire time period.

[10] However, he also points out that these results are problematic in the sense, that unemployment and cyclical parameters underlie a simultaneity bias. Does higher GDP cause decreases in unemployment or does lower unemployment cause GDP to rise. Thus the causal direction of the detected correlations can not be determined with certainty. This simultaneity makes controlling for cyclical and productivity parameters tricky.

stock-flow matching function, adding the inflow of vacancies and unemployment. These add significant information to the computation of matching efficiency, as I explained in chapter 4. Other papers which use stock-flow matching also do not consider additional cyclical effects e.g. Aranki and Löf (2008).

Thus, since I carry out a regression with Stock-Flow-Matching, I do not see a need to generally account for additional cyclical factors such as GDP or regional versions of a production measure. Nevertheless, it is not impossible that cyclical shocks have an effect on matching efficiency beyond the cyclical effect captured by unemployment and vacancies. Because the global financial crisis had such a large impact on the Swedish and global economy, I include dummy variables for the years 2008 and 2009 to capture cyclical effects in these years.

Institutional factors

Labor market institutions are often considered to influence matching efficiency (Bouvet 2012, Daly et al. 2012). These include for example the value and duration of unemployment benefits and minimum wages. However, I only focus on one country. Sweden is relatively centralized and labor market institutions such as the amount and duration of unemployment benefits are the same across all labor market regions. Changes in labor market institutions on the national level should affect all regions equally, and thus could be partly captured by the time trend if the effect diffuses over time or by year fixed effects. I try to avoid relying only on year fixed effects, since I already employ the general and region- specific time trend. Institutional changes can also be controlled for directly if the concrete reforms and corresponding variables are considered, therefore I focus on this approach.

One relevant reform in the concerned time period is the reform of disability insurance (DI) and sick pay leave, which most importantly affect older workers. In 1997 the access to disability benefits was restricted. Before that, workers above 60 had more generous access to disability benefits and labor market reasons were sufficient grounds to be granted DI. The intention of the reform was to increase employment in the age group 60 - 64 years. However, it has also been found, at least in the directly following years, that people in this age group have started to increasingly claim sick pay benefits or been unemployed instead. Sickness pay benefits are intended for shorter durations, but have been increasingly extended to longer durations (2 years and longer) (Karlström et al. 2008). Following that, sickness pay has however also been undergoing reforms, especially since the the center-conservative government came into power in 2006. The goal was to reduce the exceptionally high rates of long term sickness beneficiaries and to improve employment rates among older workers. A wide range of aspects was changed. Importantly, a time limit for receiving

sickness benefits was introduced and beneficiaries are assessed more regularly along a rehabilitation chain (OECD 2009). Consequently, since the mid-2000s the number of beneficiaries of sickness insurance above 55 years of age has dropped considerably (Spasova et al. 2016).

Both reforms, disability insurance and sick pay leave, are likely to lead to more unemployed in older age, who would have otherwise just left the work force. Such an increase in old unemployed can have an impact on matching efficiency, since this is a group which is typically assumed to face difficulties when searching for a new job. I therefore include a control capturing the number of persons above 55 in the unemployment pool. Thus, this control should account for changes in matching efficiency caused by these reforms.

Another relevant institutional change is the increased importance of labor market programs and other forms of subsidized work. According to labor market data from the Public Employment Service, in January 1998, 19 % of all registered job seekers were in a labor market program (program med aktivitetsstöd). This share decreased somewhat during the 2000s to around 10 % and has been significantly increasing since 2010. It stood at 26% in January 2016 (Arbetsförmedlingen 2022). The importance of subsidized work has changed too. For example, with the labor market reforms after 2006 "New-start jobs" (Nystarts job) were introduced, which are government subsidized jobs for persons previously on disability or sickness benefits (OECD 2009). Changes in these kind of programs can be expected to have some influence on matching efficiency, since job seekers in such a program are not required to actively search for a job to receive unemployment benefits. While openly unemployed receive unemployment benefits contingent on that they actively look for work and can be excluded from them if they decline a 'suitable' job offer, people in programs do not have so strict requirements for job search to obtain these benefits (Sianesi 2008). To prevent that this development confounds the regression it seems sensible to use openly unemployed job seekers (Öppet Arbetslösa) for my matching efficiency measure and the covariates, as these exclude job seekers in programs and only comprise individuals who are actively looking for work. Thus it excludes all variations in matching efficiency which are caused by the rise in such labor market programs. Another option is to use matching efficiency as computed with total job seekers and add a control for the share of persons in programs in the panel regression.

Composition of the working population and unemployment pool

Different socioeconomic groups exhibit different matching behaviors. Recent findings on the Swedish labor market indicate for example that unemployment is especially concentrated among persons who are young, foreign born or have low educa-

tion (Konjunkturinstitutet 2021). Other studies in the literature also identify that changes in the composition of the labor force and the unemployment pool can alter the overall matching efficiency (Barnichon and Figura 2011, Hall and Schulhofer-Wohl 2018). Additional to young people, the share of women in the labor force is assumed to decrease matching efficiency, because both groups have lower attachment to a job. There is evidence that higher shares of long term unemployed and women in the unemployment pool lower the matching efficiency (Bouvet 2012). Börsch-Supan (1991) uses similar controls for unemployment composition and finds that they have significant effects on the Beveridge curve. Generally, employed job seekers could be expected to find a job more easily, since the attribute of being unemployed is considered a negative signal by some employers (Layard et al. 2005). Concerning foreign born persons, studies have found that integrating immigrants into the labor market has been challenging in the Nordic countries because of high minimum or union negotiated wages, language barriers and high requirements of formal education (Jakobsen et al. 2019).

Education is also considered as a relevant composition variable by some studies (Petrongolo and Pissarides 2001). Education levels in the population have increased since the 1990s and labor demand tends to emphasize more specialized qualifications than before (Eklund et al. 2015). Moreover, it is possible that the overall increase in education level is correlated with the number of people with IT education, since this group only includes people with medium or high education levels. The direction of the education effect on matching efficiency is however not entirely clear. Eklund et al. (2015) finds that matching efficiency in Sweden differs by education groups, highly educated job seekers are less affected by cyclical shocks than low educated ones, so they tend to have higher matching efficiency. Coles and Smith (1996) on the other hand argue that highly educated workers can also be thought to be more specialized and less open for the different jobs, therefore decreasing matching efficiency.

Summing up, in my analysis I will include controls for old, young and female unemployed, education levels in the overall population and the number of foreign born persons in the overall population. When using matching efficiency computed by total unemployed, I will also include a control for the number of employed job seekers.

One critical remark about using controls for the composition of the unemployment pool is that the use of digital tools in recruiting might alter the job finding chances differently for different socioeconomic groups. Therefore, changes in the composition of the unemployment pool could be caused by increased IT use. In that case, controlling for the composition of the unemployment pool might bias

the estimated effect of digitization. How exactly could the digitization in recruiting change the composition of the unemployment pool? In the early years of the studied period (1998 - 2007) IT induced improvements in matching efficiency should relate mostly to the introduction of online job search and professional networks online. Young and middle-aged job seekers might be more likely to use and benefit from those than older job seekers. Assuming that digitization would benefit mostly middle aged and young job seekers, the average unemployment duration of this group would fall, while it would stay constant for old job seekers. Therefore, the share of old job seekers in the unemployment pool would rise. If this scenario would apply, it might be more plausible to measure old unemployed not in terms of the share of the unemployment pool, but as an absolute number. This way I avoid measuring an increase in the share of old unemployed, when their absolute number has actually remained constant and only overall unemployment has declined. Using absolute numbers has however the disadvantage that it also contains information on the unemployment level at the time, which can lead to other bias and which then needs to be controlled for in turn.

In the later years of the studied period, 2010 onward, algorithmic decision making and software use emerge. I do not identify a mechanism by which general software use would impact the composition of the unemployment pool. But sophisticated algorithmic decision making has the potential to be less discriminating towards groups which typically face labor market discrimination, because flaws in human judgement could theoretically be eliminated (Black and van Esch 2020). Currently however, the evidence indicates that it exhibits the same or even greater biases than human decision making because of biases in the training data (Munoz et al. 2016, Caliskan et al. 2017). Therefore, the increased use of algorithmic decision making could make it harder (or easier) to find jobs for those who exhibit characteristics labeled as unattractive, for example women and long-term unemployed. But given that my panel ends in 2016, it seems unlikely that sophisticated algorithmic decision making is already so widely used that it significantly altered the composition of the unemployment pool.

6.2 Results

In the previous section I have elaborated the functional form of my regression, which variables I add as controls and which limitations remain. Based on these findings I present different analyses and their results in this section. In the first subsection I present the results of fixed effects analysis with a time trend. In this first subsection I also elaborate summary statistics for the covariates and check whether general OLS

assumptions are fulfilled. Second, I present results of a model where the time trend is substituted with year fixed effects. Third, I run a first difference regressions to check the robustness of the results. Fourth, I add a regression with lagged explanatory variables to exclude unobserved confounders and reverse causality. Fifth, considering population size, I analyse large and small labor market regions separately.

6.2.1 Fixed Effects Regression with Time Trends

At first I will consider a fixed effects regression. Additionally to the region fixed effects μ_i, I also include a region-specific time trend $\mu_i * t$, controlling for the individual trends of matching efficiency in each region. A general time trend t is included as well, to capture the general decline of matching efficiency in Sweden which has been found in previous studies and which I observe in my data too. A drawback of including these trends is that they do not capture unobserved events in each year very well, therefore I will carry out regressions in the next section in which I include year fixed effects. Thus my first variation of the fixed effects regression will have the following form.

$$\omega_{it} = \alpha + \gamma_1 ln\ Digi_{it} + \gamma_2 ln\ C_{it} + \gamma_3 t + \gamma_4 \mu_i + \gamma_5 \mu_i * t + \epsilon_{it} \tag{17}$$

Digi is the aggregated number of individuals with an IT education who work in region i in year t. C is a vector of the control variables for compositional changes of the labor force and unemployment pool. These control variables were elaborated in the last section and are summarized below, each variable is expressed for region i in year t.

- Female Unemployed: The aggregated number of women in the unemployment pool (either open or total unemployed)
- Old Unemployed: The aggregated number of individuals in the unemployment pool who are aged 55-64 (either open or total unemployed)
- Young Unemployed: The aggregated number of individuals in the unemployment pool who are aged 18 - 24 (either open or total unemployed)
- Foreign Born: The aggregated number of individuals who were born abroad
- Low Education: The aggregated number of individuals who possess a low education level, that means persons who have 9 or 10 years of education or less (Förgymnasial utbildning, Sun2000niva levels 1 and 2)

- Medium Education: The aggregated number of individuals who possess a medium education level, that means persons who have secondary education (Gymnasial utbildning) of any duration or post secondary education (Eftergmnasial utbildning) of less than two years length (Sun2000niva 3 and 4). This variable is not included in the regression and therefore acts as the reference group for the coefficients of high and low education.

- High Education: The aggregated number of individuals who possess a high education level, that means persons who have post secondary education with more than 2 years length or who have research education (Forskarutbildning) (Sun2000niva 5 and 6)

- Year 2008, Year 2009: Dummy variables for the years 2008 and 2009 to account for the global financial crisis.

As a standard these variables are absolute values which I take the logarithm of. In some regressions the absolute number is divided by the total population of the region before the logarithm is applied, and are therefore expressed as shares of the population. Consequently the regressions are labeled as "Shares" or "Abs.".
In regressions with absolute values I always include two additional variables which I do not include in regressions with shares.

- Employment: The aggregated number of individuals who are working in a region.

- Population: The aggregated number of individuals living in a region.

Employment is included because the absolute numbers of unemployment groups entail information on the unemployment level. This control thereby avoids mixing the effects of the unemployment composition and the overall unemployment level. The second additional control is population size, which is supposed to control for changes in the absolute numbers of any of the independent variables which is purely due to changes in population size. For regressions with total unemployed I further add two more variables to control for composition changes of the pool of total job seekers.

- Employed Job seekers: The aggregated number of registered job seekers who are currently employed and do not receive government support.

- Unemployed in Programs: The aggregated number of unemployed individuals who are in labor market programs.

Summary Statistics

I will briefly consider the overall behavior of these covariates in Table 3. We can observe that female unemployed comprise a larger group than young and old unemployed. We can also observe that, in the average labor market region, most people have medium education and a similar number of people has low or high education. I also observe that there is a region with no IT specialists in one year. Since I employ logs in the regression and want to avoid missing observations, I add 1 to the number of IT specialists (Digi) for each observation before I take the logarithm. Focusing

Table 3: Summary Statistic, Independent Variables, Absolute values (Abs.)

	Mean	SD	Min	Max	N
Digitization (IT specialists)	935.7	3230.5	0	34996	1368
Old Open Unemployed	427.0	892.1	7	7272	1368
Old Total Unemployed	1507.4	2872.5	41	24141	1368
Young Open Unemployed	525.0	1024.6	6	8867	1368
Young Total Unemployed	1608.0	2784.9	20	22706	1368
Female Open Unemployed	1337.0	3195.5	13	26321	1368
Female Total Unemployed	4745.7	9463.7	62	75988	1368
Foreign Born	16337.2	55484.4	68	568548	1368
High Education	24332.2	76301.1	207	772389	1368
Medium Education	50368.8	118785.9	1133	953027	1368
Low Education	27696.5	57922.1	549	445211	1368
Employed Job Seekers	2478.7	4728.9	26	40709	1368
Unemployed in Program	1761.5	3550.6	14	33512	1368

on the compositional unemployment groups, I consider how many of the total job seekers of each group are likely to be openly unemployed. This might be relevant when interpreting potential differences between the regression with open or total unemployed. Taking the average of the shares of all FA regions in all years, young job seekers are the the most likely to be openly unemployed (29% of total young job seekers are openly unemployed). This number is at 26% for old job seekers and 23% for female job seekers (Table A18 in Appendix 1).[11] Considering which share of the unemployment pool each group makes up for: Women make up almost half of the unemployment pool (41% of the openly unemployed, 49% of the total unemployed are women). Old and young unemployed each make up around 18 % of the unemployment pool, for both open and total unemployment (See Table A19 in the Appendix). A figure presenting changes of these shares over time for Sweden as a country can be found in the appendix (A5).

[11] Please note that in the calculation of these averages, the average is taken from the shares of these numbers in all regions in all years, therefore each region is weighed equally. Therefore these numbers might not be representative for the shares on the aggregate country level, Sweden.

I also present summary statistics for the control variables when presented as population shares in the appendix (Table A20). The size ratios between the unemployment covariates are similar to the absolute values. I notice that the share of total job seekers in the population is rather high. While for example on average only 1.2% of the population in a FA region is female and openly unemployed, 5.3% is female and belongs to the group of total job seekers. This seems to be a relatively high share (keep in mind this is the rate on the total population not the working population) and confirms my earlier doubt, that the number of total job seekers might overestimate the actual number of job seekers. Considering the education levels, I can observe that on average 50% of the population in each FA region in each year has medium education, while 15% have high and 30% have low education.[12]

Residuals and Standard Errors

After estimating the regression with all control variables for the first time without adjusting standard errors, I analyze the residuals to check how well the model fits the data. The figures can be found in A6 in the appendix. Plotting the residuals over time I do not observe any clear trends which are still present in the residual's distribution. This is a good indication for the validity of the model. Besides that I also check for the heteroskedasticity of the error terms. Heteroskedasticity does not cause bias or inconsistency of the coefficients, but homoskedasticity is required for the Gauss Markov Theorem to apply and to be able to apply confidence intervals. The Preusch-Pagan Test rejects the null hypothesis that the residuals exhibit a constant variance (homoskedasticity). The heteroskedasticity can also be observed in the scatter plot of residuals and fitted values. As expected, I also detect serial correlation of the residuals by running a simple auto regression (Table A21). Lastly, I check whether the residuals are normally distributed. Plotting the residuals in a histogram against the normal distribution I find that the the normal distribution fits the residuals quite well. However, when I run formal normality tests, such as the Jarque-Berra or the Skewness-Kurtosis Test they reject the hypothesis of a normal distribution. Normality is not required for the Gauss Markov Theorem, but in order to use the standard errors for hypothesis tests and to compute t-statistics the population error needs to be independent of the explanatory variables and to be

[12] Comparing these numbers to the absolute values, the share of highly educated seems lower than the average absolute number of highly educated. When using absolute numbers for calculating an average, regions with a high population weigh relatively more. When computing average from population shares, all regions are weighed equally. Thus, this difference is an indicator, that on average, regions with a large population have a more highly educated population than regions with small population sizes.

normally distributed (Wooldridge 2018).

As I explained earlier, I cluster standard errors at the region level to control for serial correlation in the FE regression. Given these results, additionally I also employ standard errors robust to heteroskedasticity. Taking into account the results of the normality test, I also employ regressions with clustered parametric bootstrapped standard errors as a robustness check. The bootstrap method is a way to compute standard errors via a resampling method, which has the advantage that it does not make any assumption about the distribution of the error terms (Wooldridge 2018). I only use it as a robustness check, since other authors with similar papers and especially Higashi (2020), whose approach to regressing matching efficiency I follow, do not use this method. When employing the bootstrap method I cluster on region, which means each resampling draw is a sample of the clusters. Therefore the size of the drawn samples is the same size as the number of clusters. I employ 50 repetitions of the resampling process.

Regression Results

Now I turn to the actual results of the fixed effects regression with a time trend which can be found in Table 4. In the first column I start with the simplest form of my the fixed effects regression with no controls besides the region dummmies, $\omega_{it} = \alpha + \gamma_1 ln\ Digi + \gamma_4 \mu_i + \epsilon_{it}$. The R^2 already has a rather high value of 0.94. In dummy regressions high R^2 are not uncommon, due to the high numbers of variables included which account for unobserved effects. In this regression digitization has a significant and negative effect on matching efficiency. This result makes a lot of sense, since matching efficiency in Sweden has been deteriorating over time, while digitization has been increasing over time. Therefore the digitization coefficient is likely to be highly confounded with the time trend. This hypothesis is indeed confirmed by the next column, where I include a general time trend, region specific time trends and dummies for the years 2008 and 2009, to account for the global financial crisis. The effect of digitization then jumps from -0.39 to 0.22 and maintains statistical significance, while we can see that the time trend and the year dummies have a negative coefficient as expected. Adding composition variables for education levels in the population, the coefficient declines. Given that R^2 is barely changing, this implies that digitization is correlated with changes in the population's education composition. Both low and high education have positive coefficients, indicating that both groups have a higher matching efficiency than the group with medium education levels. The coefficient of low education is statistically significant and larger (1.1) than high education (0.3). This result is not in line with other studies on the Swedish labor market, which observe that lower

Table 4: FE Regression, Time Trend, Open Unemployment, Abs.

	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi	-0.3903***	0.2198***	0.1858***	0.1572***	0.1356***	0.1356***
	(0.0505)	(0.0376)	(0.0384)	(0.0332)	(0.0291)	(0.0270)
Year 2008		-0.0615***	-0.0856***	-0.1153***	-0.1161***	-0.1161***
		(0.0114)	(0.0115)	(0.0147)	(0.0141)	(0.0147)
Year 2009		-0.1940***	-0.2161***	-0.2237***	-0.2194***	-0.2194***
		(0.0116)	(0.0123)	(0.0126)	(0.0130)	(0.0104)
Time Trend		-0.0588***	-0.0231**	0.0304**	0.0167	0.0167
		(0.0015)	(0.0112)	(0.0133)	(0.0162)	(0.0157)
ln High Edu			0.3029	-0.0428	-0.1788	-0.1788
			(0.1839)	(0.1796)	(0.1845)	(0.1697)
ln Low Edu			1.1017***	1.6475***	1.5369***	1.5369***
			(0.2893)	(0.2611)	(0.3764)	(0.3300)
ln Old Unemp.				0.0116	0.0144	0.0144
				(0.0199)	(0.0212)	(0.0153)
ln Young Unemp.				0.0302	0.0342	0.0342
				(0.0267)	(0.0280)	(0.0317)
ln Female Unemp.				-0.1163***	-0.1115***	-0.1115***
				(0.0331)	(0.0361)	(0.0333)
ln Foreign born				-0.5569***	-0.4776***	-0.4776***
				(0.1020)	(0.1157)	(0.1072)
ln Population					-0.9678	-0.9678
					(0.6942)	(0.6900)
ln Employment					0.5501**	0.5501**
					(0.2341)	(0.2168)
Constant	-0.2229*	116.4830***	35.5834	-70.2819**	-38.0632	-38.0632
	(0.1165)	(2.8615)	(24.1961)	(27.4339)	(33.5155)	(31.4582)
Region Time Trend	No	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9416	0.9852	0.9857	0.9874	0.9876	0.9876
Number of observations	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with openly unemployed and seasonal fixed effects. Control variables are absolute values. Standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

skilled persons are less likely to find work (Konjunkturinstitutet 2021). A potential reason for the result could however be that highly educated persons are more likely to be searching for very specialized work, while low educated are more flexible and willing to take on a variety of jobs which leads them to be unemployed for shorter time periods (Coles and Smith 1996). Adding controls for the composition of the unemployment pool and the number of foreign born people further decreases the digitization coefficient, implying a correlation between digitization and these variables. The foreign born population exhibits a negative correlation coefficient, which is in line with the expectation. The numbers of old and young unemployed are insignificant, and female unemployed have a negative correlation coefficient. These absolute numbers of the unemployment groups may also contain information on the current unemployment level which is closely linked to economic cycles. If there is some cyclical variation in matching efficiency left, then these coefficients could be confounded with cyclical changes. Hence, I also include a control for employment, which controls for such cyclical variations at least partly. Lastly, I add a control for population size, which controls for variations in the control variables which are purely due to shifts in population size. Including these additional controls alters the other correlation coefficients only slightly. The fact that the coefficient for employment is positive and statistically significant can be seen as an indication that there is indeed some cyclical variation left in the computed matching efficiency. Since the coefficient of employment is positive it can be inferred, that matching efficiency is better in an economic boom than in a crisis. Including this full set of controls, the time trend looses significance, indicating that the controls jointly can explain at least parts of the overall downward trend of matching efficiency. Including the full set of compositional variables reduces the coefficient of digitization but it is still positive, statistically significant and has a moderate magnitude with a coefficient of 0.1356. As a robustness check I run the same regression with bootstrapped standard errors. The bootstrapped standard errors are very similar to the robust standard errors and do not change the significance levels of any explanatory variable.

Robustness checks with different data types

To check the robustness of the results I employ the same regression with variations in the used data, which can be found in Table 5. A list with an exact description of each variable in the different specifications, as well as the data source can be found in Appendix 2. The three regressions on the left are carried out with matching efficiency as computed with openly unemployed, the three regressions on the right use matching efficiency as computed from total job seekers. As I explained in chapter 4, when openly unemployed are used to compute matching efficiency, matching

efficiency is overestimated, when total unemployed are used, matching efficiency is underestimated. The two measures represent the minimum and maximum representation of actual matching efficiency. Each regression will be discussed in detail below. Full regression tables for each regression separately can be moreover found in the appendix (see Tables A22, A23, A24). The first column entails the regression

Table 5: FE Regression, Time Trends, Comparison of Specifications

	Openly Unemployed			Total Job Seekers		
	Shares	Absolute	Absolute	Shares	Absolute	Absolute
	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi	0.0868***	0.1596***	0.1356***	0.0546*	0.1200***	0.0962***
	(0.0318)	(0.0349)	(0.0291)	(0.0280)	(0.0336)	(0.0307)
ln High Edu	-0.1987	0.0731	-0.1788	0.2351	0.4203**	0.0924
	(0.1763)	(0.1854)	(0.1845)	(0.1795)	(0.1641)	(0.1818)
ln Low Edu	1.5189***	1.7771***	1.5369***	1.4174***	1.5801***	1.2005***
	(0.2908)	(0.3006)	(0.3764)	(0.2460)	(0.2355)	(0.2854)
ln Foreign born	-0.2558**	-0.4728***	-0.4776***	-0.2579***	-0.4668***	-0.3447***
	(0.1018)	(0.1118)	(0.1157)	(0.0956)	(0.1077)	(0.1080)
ln Old Unemp.	0.1102***		0.0144	0.2305***		0.1403*
	(0.0234)		(0.0212)	(0.0613)		(0.0761)
ln Young Unemp.	0.1670***		0.0342	0.1720***		0.1111***
	(0.0314)		(0.0280)	(0.0381)		(0.0393)
ln Female Unemp.	0.1328**		-0.1115***	0.0278		-0.2925***
	(0.0661)		(0.0361)	(0.1065)		(0.0827)
ln Population		-1.4423**	-0.9678		-1.6484**	-1.7043**
		(0.6241)	(0.6942)		(0.6280)	(0.7369)
ln Employment			0.5501**			0.6155***
			(0.2341)			(0.1809)
ln Employed Job Seekers				0.1134**	0.0469	0.0881**
				(0.0444)	(0.0319)	(0.0416)
ln Unemp. in Programs				-0.0509**	-0.0649***	-0.0600**
				(0.0232)	(0.0148)	(0.0233)
Year 2008	-0.1149***	-0.1025***	-0.1161***	-0.0693***	-0.0748***	-0.0600***
	(0.0104)	(0.0110)	(0.0141)	(0.0100)	(0.0117)	(0.0124)
Year 2009	-0.2225***	-0.2342***	-0.2194***	-0.1718***	-0.1675***	-0.1706***
	(0.0135)	(0.0120)	(0.0130)	(0.0129)	(0.0120)	(0.0127)
Time Trend	0.0035	0.0206*	0.0167	0.0258***	0.0393***	0.0199
	(0.0114)	(0.0122)	(0.0162)	(0.0090)	(0.0110)	(0.0136)
Constant	-6.3831	-41.5057	-38.0632	-49.4751***	-76.9252***	-38.0303
	(23.0566)	(25.7828)	(33.5155)	(18.1702)	(24.6732)	(30.3148)
Region Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Robust
R-squared	0.9880	0.9872	0.9876	0.9226	0.9181	0.9220
Number of observations	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with either total or open unemployed and seasonal fixed effects. Control variables are either absolute values (Absolute) or expressed as shares of the population or unemployment pool (Shares). A description of all variables can be found in Appendix 2. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

with explanatory variables expressed as shares of the population. In this regression the variables of unemployment groups are divided by the total number of unemployed and the other variables are divided by population size. Column 3 in contrast depicts the previous regression with absolute numbers. In the regression with absolute values, all variables refer to the absolute number of people in region i in year t with the certain characteristic, e.g. with digital skills, who were born abroad, who are female and unemployed, etc.

Considering this regression in column 1, the coefficient of Digi is still significant but reduced in size (0.087). The results of other controls are very similar to column 3 for most variables but the compositional unemployment variables. In contrast to the previous regression, the coefficient of female, old and young unemployed are positive. These composition coefficients are generally less in line with the expected behavior. Old unemployed are expected to be less able to find new work and young and female persons are often assumed to be less attached to the labor market. The question remains why the coefficients are different from the absolute value regression. Shares values have the advantage that they do not entail direct information about the unemployment level and automatically excludes variation due to changes in population size. However, changes in a group's share of the unemployment pool, can also just be driven by variations in the size of the unemployment pool due to cyclical or structural reasons. In this regard I mentioned earlier, that using absolute values for the unemployment groups might be more valid if digitization would directly cause changes in the unemployment pool composition.[13] Next I consider each unemployment group individually. The absolute number of female unemployment is negatively related to matching efficiency, but the share of women in the unemployment pool is positively correlated. The later could be because the share of women in the unemployment pool decreases in economic downturns because the share of male unemployed is higher in economic downturns. Since matching efficiency is worse in economic downturns the positive female coefficients indicates that the female share in the unemployment pool could be mainly approximating economic cycles.

The coefficients of old and young unemployed are insignificant when using absolute numbers, but positive when using their share in the unemployment pool. Looking at the changes in aggregate unemployment numbers in Sweden (see Figure A5 in the appendix), the absolute numbers of old and young unemployed seem very stable over time and barely follow the cyclical changes of the overall unemployment

[13] For example, because it improves matching efficiency for some groups more than others and thereby changes the shares of an unemployment group in the unemployment pool, without changing the absolute number of this group.

numbers. That means that there is little variation over time in these variables. There could be different reasons why these absolute numbers exhibit less cyclical variation. Possibly old unemployed are more likely to get granted early retirement options when the labor market situation is worse. Even though labor market reasons are not officially a sufficient reason for that, civil servants might be more benevolent in these situations. Young persons might be more likely to extend their education when the labor market situation is worse. Thus, old and young persons could be less affected by cyclical shocks than the prime age working population. In turn, that means that the share of these groups in the unemployment pool would be mainly driven by changes in the size of the unemployment pool and not by actual changes in the absolute numbers. In that case, the positive coefficients of the share variables could be caused by a confounding with changes in the size of the unemployment pool, these in turn are most likely driven by cyclical effects.

In column 2, I run a regression with absolute numbers without any compositional unemployment controls. In that case the digitization coefficient is 0.1596. When running the regression with share values without the compositional unemployment controls, the coefficient is 0.1435 (see Table A22 in the Appendix). Thus the main difference in the digitization coefficient between regression 1 and regression 3 can be explained by the behavior of the compositional unemployment variables. I explained why I think that the results in the share regression are more likely to be confounded than the results in the absolute value regression.

In column 4 to 6 I run the same regressions as before, but use matching efficiency (and unemployment control variables) as computed with the total job seekers.[14] In this case I further add controls for the number of job seekers who are in labor market programs and for the number of employed job seekers (on-the-job job seekers) in the total unemployment pool, because I suspect that compositional changes of these groups could also affect matching efficiency. In the regression with openly unemployed such changes are excluded from the dependent variable by design. Column 6 includes a regression with absolute values, comparable to column 3. The coefficient of digitization is smaller than in column 3, when openly unemployed are used (0.0962 vs. 0.1356), but it is still statistically significant. The coefficients of the control variables are similar to the regression with openly unemployed. A difference is that the coefficient of young unemployed is positive and significant, but here the results for the regression with shares and with absolute values are more in line. As

[14] Summary statistics of between and within variation for both types of matching efficiency were displayed in chapter 4. I observed that within variation is larger for the open unemployed matching efficiency than the total one. In Tables A23 and A24 in the appendix I also display the full regression tables for the regressions with total job seeker matching efficiency.

I mentioned in the summary statistics, young unemployed are the most likely to be openly unemployed which can also lead to differences in the correlation coefficients when using either open or total unemployed. The number of unemployed in programs has a negative correlation coefficient and the number of job seekers who are still employed has a positive one. These coefficients are in line with the expectation. Turning to the results in column 4, which measure all control variables as shares, it can be observed that the digitization effect is smaller in magnitude (0.054) and statistical significance. Other control variables also vary slightly, but are not fundamentally different from column 1.

Considering all specifications with full controls, the digitization coefficient varies between magnitudes of 0.05 and 0.1356 and is always at least marginally significant. Therefore, these results imply that there could be some effect of digitization on matching efficiency. However, compared to other controls, the correlation coefficient is relatively small, shares of education levels and foreign born persons in population seem to have a larger impact on matching efficiency. Thus, although digitization seems to has some influence, other labor market conditions seem to be more relevant.

6.2.2 Fixed Effects Regression with Year Fixed Effects

As a next robustness check I include year fixed effects.

$$\omega_{it} = \alpha + \gamma_1 ln\ Digi_{it} + \gamma_2 ln\ C_{it} + \gamma_3 \lambda_t + \gamma_4 \mu_i + \epsilon_{it} \tag{18}$$

Year fixed effects capture unobserved effects on matching efficiency separately for each year. Therefore, they allow unobserved region-constant time-varying effects to vary for each year and capture unobserved effects more granular than a time trend. The time trend considers a constant unobserved effect for all time periods. If an unobserved event has a particular effect only in a certain year, the trend is less precise at capturing this than year dummies are. Therefore, a regression with year fixed effects has the advantage of better accounting for unobserved time events which affect all regions evenly e.g. economic crisis, labor market reforms, etc. But only accounting for year fixed effects includes the risk of not accounting for unobserved region specific trends. As I explained in section 6.1.2 region specific time trends are however relevant, since they capture region specific developments, such as structural economic shift.

Regarding this point, an alternative could be to combine all of these measures together, region dummies, year dummies, a time trend and region-specific time trend. But, this carves out a lot of the variation. The only variation left then, is the year-region variation which exceeds the average region and average year effect

and which is also exceeding the average time trend and the region specific-time trend. That is very little variation to find any significant correlation coefficients for the explanatory variables. The results of such a regression with time trends and year fixed effects can be found in the appendix for open and total unemployment (Tables A26 and A27). In these regressions the digitization effect is statistically and economically insignificant, but some other control variables turn insignificant too and they are not robust between the two regressions. Only high education and female unemployed seem to have a robust effect in this setting. These results confirm that there is very little variation left after applying this functional form.

Therefore, I will now focus on a regression with only region and time fixed effects and exclude the time trend. The results of the regression are displayed in Table 6. After running the regression for the first time without adjusted standard errors, I again check the distribution, normality and heteroskedasticity of the residuals. The results can be found in the appendix (Figure A7, Table A25) but are essentially the same as for the Fixed Effects Regression with Time Trends and I proceed with clustering robust standard errors. Considering the results it is first observed that the baseline correlation between matching efficiency and digitization is much smaller. In the previous FE Regression the coefficient was 0.219 when only time trends and region dummies were used as controls. With year and region fixed effects the baseline correlation is only 0.097 and decreases further with adding more controls. With additional controls the Digi coefficient becomes economically and statistically insignificant.

When considering the difference between the regressions with time trends and the regression with year fixed effects, it can be observed, that in the time trend regression, digitization and matching efficiency exhibit a relatively robust statistically significant correlation, with a correlation coefficient of moderate magnitude. With year fixed effects, these coefficients are dramatically reduced. This could be because there are indeed unobserved events in different years which are correlated with matching efficiency and digitization. Another reason could be that there is too little variation in the digitization variable. If increases in the level of IT skills between different regions are very similar within years, then the correlation between digitization increases and matching efficiency might be partly taken up by the year fixed effect instead. Furthermore, a confounding with region specific trends, such as structural economic change can not be excluded.

Next, I consider the results of the other explanatory variables. A relevant difference to the regression with time trends is that both education coefficients are negative. But in line with the previous results, high education is worse for matching efficiency than low education, since the coefficient of high education is statistically

Table 6: FE Regression, Year Fixed Effects, Open Unemployed, Abs.

	(1)	(2)	(3)	(4)	(5)
ln Digi	0.0970***	0.0376	0.0423	0.0218	0.0218
	(0.0349)	(0.0363)	(0.0352)	(0.0292)	(0.0276)
ln Old Unemployed	0.0013	0.0024	0.0042	-0.0046	-0.0046
	(0.0331)	(0.0311)	(0.0328)	(0.0325)	(0.0323)
ln Young Unemployed	0.0663*	0.0021	0.0057	0.0012	0.0012
	(0.0361)	(0.0378)	(0.0374)	(0.0372)	(0.0402)
ln Female Unemployed	-0.1791***	-0.1634***	-0.1830***	-0.1483**	-0.1483***
	(0.0503)	(0.0553)	(0.0547)	(0.0565)	(0.0572)
ln High Edu		-0.3223	-0.5389**	-0.4594**	-0.4594**
		(0.2180)	(0.2283)	(0.2120)	(0.2317)
ln Low Edu		0.8857***	0.2196	-0.0881	-0.0881
		(0.2301)	(0.4203)	(0.3863)	(0.4315)
ln Population			1.3820**	1.6571***	1.6571***
			(0.6010)	(0.5519)	(0.6431)
ln Employment			-0.4353*	-0.4933*	-0.4933**
			(0.2605)	(0.2500)	(0.2162)
ln Foreign born				-0.1782**	-0.1782**
				(0.0807)	(0.0906)
Constant	0.2184	-4.6583**	-6.6303***	-5.3551**	-5.3551*
	(0.2846)	(2.1103)	(2.4487)	(2.3741)	(2.7550)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9849	0.9861	0.9864	0.9867	0.9867
Number of observations	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are share values. All standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

significant while low education is not. The negative sign of the coefficients could potentially be due to a spurious correlation with time, because the share of highly educated persons has been increasing over time and matching efficiency has been decreasing over time (see Figure A4 in Appendix 1). The coefficients of the compositional unemployment variables are in line with the previous regressions and seem to be less relevant. The significance of foreign born persons is much more robust in this specification than with the time trend, but the direction and magnitude of the effect is the same. Since the foreigner effect is negative and the digitization effect is positive, this implies that digitization changes are negatively correlated with changes in the number of foreign born persons, when year fixed effects are controlled for. The coefficient of the foreign born population has been significant in the regressions with time trends too, but had a smaller and less significant coefficient.[15] Generally, the number of foreign born persons in Sweden has been increasing steeply and also exhibits variations in the rate of increase between different years (Figure A3, Appendix). Therefore it is not surprising that it does seem to have a robust effect on matching efficiency.

As in the previous subsection, I check the robustness of the regression by using variations in the data used (Tables A28, A29, A30 in Appendix 1). These show a similar pattern, although in these regressions the correlation coefficient of Digi maintains a slightly larger significance level after controls are added. When using population shares for the variables, the digitization effect remains marginally significant after compositional controls and education levels are added, but turns insignificant once all controls are added. The same pattern can be observed in the regression with total job seekers. The coefficient of Digi always stays positive and the coefficients of the controls are relatively stable across the different specifications. The control for foreign born persons seems to be especially relevant and also substantially decreases the digitization effect in these regressions too.

6.2.3 First Difference Regression

As a robustness check, I also estimate results with a first difference approach, which is another method to eliminate region specific unobserved characteristics. The basis for the functional form of the first difference regression are the fixed effects regressions from the previous sections. First I take difference of the regression with the time trend.

[15] In the time trend regression, including the coefficient of foreigners decreased the coefficient of digitization only slightly.

$$\Delta\omega_{it} = \Delta\alpha + \Delta\gamma_1 ln\ Digi_{it} + \Delta\gamma_2 ln\ C_{it} + \Delta\gamma_3 t + \Delta\gamma_4\mu_i + \Delta\gamma_5\mu_i * t + \Delta\epsilon_{it} \quad (19)$$

Assuming Δ captures a one period change between year t and year t-1, we can eliminate some variables from this equation. α and μ_i will drop out, since they are constant over time. The time trend will be reduced to a constant since $\gamma_3 * (t - (t + 1)) = \gamma_3 * 1$, following the same pattern the interaction term will be reduced to $\gamma_5\mu_i$. Thus we obtain:

$$\Delta\omega_{it} = \gamma_1\Delta ln\ Digi_{it} + \gamma_2\Delta ln\ C_{it} + \gamma_3 + \gamma_5\mu_i \quad (20)$$

The region fixed effects in this regression represent region specific time trends. Differencing also eliminates the level of year-varying unobserved effects. Adding an intercept for each year would measure the increases in year-unobserved effects, but I do not see a methodological necessity for that.

I first run a first difference regression using variables based on openly unemployed and absolute values, presented in Table 7. The R^2 is much lower in the FD regression, but this is to be expected, given that differencing takes away a part of the information of each variable. The digitization coefficient is very similar in magnitude to the fixed effects regression with the time trend (around 0.08 with all controls), but the standard errors are larger and thus the effect is mostly statistically insignificant. When employing bootstrapped standard errors there is a marginal statistical significance level. The coefficients of other controls are roughly in line with previous results. The effects of education and foreign born are confirmed, as well as the time trend (constant). Young unemployed have a negative sign, which is a result different to previous regressions, but the coefficient is not very large in magnitude.

I also run the regression without the region dummies and constant, that means without time trends (Table A31 in the appendix). This yields very similar results but the digitization coefficient maintains marginal significance with the full set of controls. In the appendix (Table A32) I also provide a comparison of the regression results with different data types, for open unemployed vs. total job seekers and absolute vs. share values. With full controls, the magnitude of the digitization coefficient is between 0.053 and 0.0915 and mostly statistically insignificant. Therefore, the regressions with first differences are in agreement with the results from the fixed effects regressions. There seems to be some correlation between digitization on matching efficiency, but it has low statistical significance levels and is of moderate magnitude. Compared to other control variables, digitization seems to be of minor importance.

Table 7: First Differences, Time Trend, Open Unemployed, Absolute Values

	(1)	(2)	(3)	(4)	(5)	(6)
Δln Digi	0.0937	0.0991	0.1022*	0.0843	0.0844	0.0844**
	(0.0574)	(0.0608)	(0.0580)	(0.0520)	(0.0526)	(0.0415)
ΔYear 2008	-0.1207***	-0.1208***	-0.1221***	-0.1453***	-0.1534***	-0.1534***
	(0.0082)	(0.0082)	(0.0080)	(0.0083)	(0.0083)	(0.0088)
ΔYear 2009	-0.2353***	-0.2352***	-0.2381***	-0.2182***	-0.2335***	-0.2335***
	(0.0103)	(0.0103)	(0.0101)	(0.0101)	(0.0107)	(0.0089)
Δln High Edu		-0.0996	-0.0359	-0.0885	0.0480	0.0480
		(0.1420)	(0.1402)	(0.1325)	(0.1495)	(0.1465)
Δln Low Edu		0.0460	0.2902	0.5687***	0.9467***	0.9467***
		(0.2192)	(0.2078)	(0.2034)	(0.2349)	(0.2158)
Δln Foreign born			-0.4155***	-0.4278***	-0.3058***	-0.3058***
			(0.0937)	(0.0915)	(0.0971)	(0.0949)
Δln Old Unemployed				0.0013	-0.0085	-0.0085
				(0.0144)	(0.0142)	(0.0114)
Δln Young Unemployed				-0.0563***	-0.0563***	-0.0563***
				(0.0182)	(0.0191)	(0.0165)
Δln Female Unemployed				-0.0816***	-0.0680**	-0.0680***
				(0.0266)	(0.0269)	(0.0225)
Δln Employment					-0.0496	-0.0496
					(0.1348)	(0.1385)
Δln Population					-1.6746***	-1.6746***
					(0.4273)	(0.4247)
Constant	-0.0509***	-0.0473***	-0.0235**	-0.0160*	-0.0235**	-0.0235**
	(0.0035)	(0.0078)	(0.0095)	(0.0092)	(0.0092)	(0.0112)
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.3588	0.3592	0.3748	0.4036	0.4105	0.4105
Number of observations	1296	1296	1296	1296	1296	1296

Note: The independent variable is matching efficiency, computed as the residual of matching function open unemployed and seasonal fixed effects. Control variables are share values and absolute. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Multi-year Differences, Openly Unemployed, Absolute

Difference period	1-year	2-year	3-year	4-year
Δln Digi	0.0900	0.0431	0.0209	0.0123
	(0.0544)	(0.0687)	(0.0690)	(0.0682)
Δln High Edu	-0.4956***	-0.3498	-0.1962	-0.0812
	(0.1813)	(0.2530)	(0.2952)	(0.3489)
Δln Low Edu	-0.2498	-0.0093	0.0717	-0.1784
	(0.2812)	(0.4213)	(0.5233)	(0.6408)
Δln Old Unemployed	0.0209	0.0672**	0.0750**	0.0516
	(0.0174)	(0.0288)	(0.0359)	(0.0386)
Δln Young Unemployed	-0.1385***	-0.1986***	-0.2076***	-0.2031***
	(0.0210)	(0.0233)	(0.0281)	(0.0326)
Δln Female Unemployed	-0.0555*	-0.1002**	-0.1361***	-0.1486**
	(0.0298)	(0.0410)	(0.0506)	(0.0571)
Δln Foreign born	-0.5186***	-0.1658	-0.1235	-0.1173
	(0.1207)	(0.2247)	(0.2470)	(0.2523)
Δln Emplyoment	0.2246	-0.2261	-0.3055*	-0.2461
	(0.1586)	(0.1438)	(0.1663)	(0.1915)
Δln Population	2.2674***	1.8239***	1.7952**	1.8652**
	(0.4113)	(0.5916)	1.7952**	(0.8075)
Constant	-0.0170	0.0063***	0.0158***	0.0091
	(0.0123)	(0.0020)	(0.0058)	(0.0067)
Region Dummy	Yes	Yes	Yes	Yes
R-squared	0.1453	0.1907	0.2348	0.2743
Number of observations	1296	1224	1152	1080

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Multiyear Differences

As another sensitivity check I run a regression with differences over longer periods, in Table 8 a regression with difference periods spanning between 1 and 4 year-differences can be found. In Appendix 2 I also present results and discussion for even longer difference periods. Like before, the digitization coefficient is insignificant when all controls are included. The magnitude of the coefficient gradually declines with increasing difference periods, but it stays positive for lag periods 1-4. The results indicate that there does not seem to be a measurable effect over longer time periods, but it confirms that the direction of the effect would be more likely to be positive than negative. Considering the behavior of the other controls, the unemployment composition is interestingly very significant compared to the other controls. Possibly changes in unemployment composition are more likely to have an effect over longer time periods.

6.2.4 Regression with Lagged Variables

To interpret the correlation coefficient of digitization as the causal effect on matching efficiency, it must be excluded that there are unobserved events which affect both, digitization level and matching efficiency. Furthermore, it must be excluded that the correlation is due to reversed causality, that means that matching efficiency itself does not cause changes in digitization level. Both cases, reversed causality or unobserved events, would lead to a correlation between $ln\ Digi_{it}$ and the error term ϵ_{it}. In order to exclude such events I introduce lagged explanatory variables. I instrument the number of digital specialists in period t with the number of digital specialists in period t-n.

$$\omega_{it} = \alpha + \gamma_1\ ln\ Digi_{i,t-n} + \gamma_2\ ln\ C_{i,t-n} + \gamma_3\lambda_t + \gamma_4\mu_i + \epsilon_{it} \tag{21}$$

In order to use the lagged term as an instrument, it must be assumed that 1) $ln\ Digi_{it}$ and $ln\ Digi_{it-n}$ are correlated and 2) that $ln\ Digi_{t-n}$ does not affect ω_{it}, through channels other than $ln\ Digi_{it}$. The second point implies that $ln\ Digi_{t-n}$ and ϵ_{it} are uncorrelated. However, if $ln\ Digi_{it}$ and ϵ_{it} are indeed correlated because of reversed causality or an unobserved effect, then $ln\ Digi_{it-n}$ and ϵ_{it-n} are correlated too. It is also plausible to assume that the error term is an auto-regressive process, that means that ϵ_{it} and ϵ_{it-n} are correlated over a certain number of periods n. In a fixed effects panel regression serial correlation of the error term is expected. As long as the error terms ϵ_{it} and ϵ_{it-n} are correlated, so will be $ln\ Digi_{it-n}$ and ϵ_{it}, and a lagged variable does not solve the issue (Wooldridge 2018). Therefore, I check the auto-correlation of the error terms and choose a value for n at which ϵ_{it-n} and ϵ_{it} are not correlated any more. In a functional form with year fixed effects this

applies for four lags, in a functional form with a time trend it is two lags (Appendix Tables A33 to A36). The digitization variable is still auto-correlated for these lag periods (Appendix Table A37). Table 9 present the result of the lagged variables

Table 9: Lagged Values: FE Regression, Year FE, Open Unemp., Abs.

	(1)	(2)	(3)	(4)	(5)
ln Digi $_{t-4}$	0.0919**	0.0866**	0.0668*	0.0091	0.0070
	(0.0412)	(0.0387)	(0.0380)	(0.0382)	(0.0364)
ln Old Unemp $_{t-4}$		-0.0380	-0.0561**	-0.0458*	-0.0404
		(0.0293)	(0.0254)	(0.0272)	(0.0277)
ln Young Unemp $_{t-4}$		0.1148**	0.0922*	0.0180	0.0160
		(0.0565)	(0.0531)	(0.0530)	(0.0490)
ln Female Unemp $_{t-4}$		0.0154	0.0896	0.0983*	0.0660
		(0.0684)	(0.0539)	(0.0576)	(0.0557)
ln Foreign born$_{t-4}$			-0.2870***	-0.2135**	-0.2769***
			(0.0717)	(0.0823)	(0.0762)
ln High Edu $_{t-4}$				0.0482	-0.3084
				(0.2432)	(0.2355)
ln Low Edu $_{t-4}$				0.9172***	-0.7223
				(0.3004)	(0.5182)
ln Employment $_{t-4}$					-0.7162***
					(0.2523)
ln Population $_{t-4}$					2.9391***
					(0.6349)
Constant	-0.5219**	-0.9922***	1.1528	-7.7555***	-12.2447***
	(0.2087)	(0.3403)	(0.7149)	(2.6402)	(2.6275)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
R-squared	0.9853	0.9858	0.9866	0.9880	0.9888
Number of observations	1080	1080	1080	1080	1080

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

regression with year fixed effects. Interestingly, the results with the lagged variable are remarkably similar to the results with the non-lagged value for the regression with fixed year effects. While the correlation coefficient also turns insignificant once all controls are included, it shows some small statistical significance and a moderate magnitude when not all controls are included. This pattern is stable when using a regression with trends instead of year fixed effects. Robustness checks with total job seekers indicate similar patterns with lower statistical significance (Appendix Tables A38 to A40). Considering the other control variables, foreign born persons and education levels seem to be of remarkable importance again and the direction of the effects are the same. The coefficients of the composition groups are mostly very small and statistically insignificant.

Thus, this analysis rather confirms that there seems to be some correlation be-

tween digitization and matching efficiency, but it is not so significant that it is maintained when other controls are added. Other variables seem to be better at explaining changes in matching efficiency than digitization.

6.2.5 Analysing Urban and Rural Regions Separately

As I mentioned in chapter 3, there are large differences between the different labor market regions in Sweden in terms of population size. The largest region had on average 1.9 million inhabitants over the time studied, while the smallest had on average 2300 inhabitants. Therefore, I am interested in how the results change when I exclude either the very large or the very small regions. Focusing on these two groups might further improve the understanding of the relationship of digitization and matching efficiency. Moreover, excluding very small labor market regions improves the external validity of the results with respect to other European countries. Very small labor market regions are a particularity of Sweden and possibly some other Nordic countries, but are rather uncommon in other parts of Europe.

To obtain reliable results I want the reduced sample to still contain at least 42 regions. Thereby I avoid methodological problems, e.g. such as how to deal with serial correlation in the panel setting when there are very few regions to cluster on (Angrist and Pischke 2008). I rank labor market regions by their average population size between 1998 and 2016. Firstly, to focus on larger regions, I exclude the 30 smallest regions from the sample, which means that the remaining 42 regions all had an average population above almost 20 000. Secondly, to focus on smaller regions, I exclude the 30 largest regions from the full sample, which means that the remaining regions all have a population below 45 600.

In Table 10 I present the results for large and small regions separately using openly unemployed for matching efficiency. The first column includes a functional form with time trends, the second one includes year fixed effects. In the regression with time trends, the digitization coefficients are statistically significant in both, large and small regions, and slightly larger in magnitude in large regions. Using year fixed effects, the correlation coefficient is almost insignificant in both large and small regions, although it is slightly larger in small regions.

As a robustness check I run the same regressions with total unemployed (Table A41), which confirms the same pattern. The marginally significant coefficient of digitization in the regression with year fixed effects becomes however insignificant in all robustness regressions. The difference between large and small regions is even larger in the time trend regression with total job seekers. Using population shares instead of absolute values confirms the same pattern too: There is barely any

Table 10: Large and Small Regions, Openly Unemployed, Abs.

	42 Largest Regions		42 Smallest Regions	
	Trend	Year FE	Trend	Year FE
ln Digi	0.1241**	-0.0567	0.1081***	0.0468*
	(0.0537)	(0.0493)	(0.0287)	(0.0264)
ln Old Unemployed	-0.0074	-0.0515	0.0046	-0.0035
	(0.0207)	(0.0630)	(0.0260)	(0.0279)
ln Young Unemployed	-0.0013	-0.0222	0.0220	0.0158
	(0.0356)	(0.0680)	(0.0322)	(0.0375)
ln Female Unemployed	-0.1078**	-0.1308	-0.0949**	-0.1732***
	(0.0453)	(0.0827)	(0.0429)	(0.0547)
ln High Edu	-0.4906***	-0.6978**	0.4767	-0.2658
	(0.1507)	(0.3321)	(0.3214)	(0.2909)
ln Low Edu	2.2668***	0.1066	1.1158**	0.1512
	(0.3341)	(0.5497)	(0.4865)	(0.5003)
ln Foreign born	-0.9092***	-0.1385	-0.4249***	-0.1693*
	(0.1363)	(0.1346)	(0.1400)	(0.0841)
ln Employment	0.0689	-1.1992***	0.5690**	-0.3266
	(0.3613)	(0.3918)	(0.2417)	(0.2907)
ln Population	-0.2360	3.0208***	-1.3163	0.1647
	(0.7603)	(0.8555)	(0.9085)	(0.8763)
Time Trend	0.0468***		-0.0180	
	(0.0092)		(0.0189)	
Year 2008	-0.1548***		-0.1062***	
	(0.0133)		(0.0185)	
Year 2009	-0.2384***		-0.2033***	
	(0.0149)		(0.0200)	
Constant	-101.6675***	-12.0944**	33.0056	3.2984
	(21.0258)	(4.5045)	(39.1308)	(3.5801)
Region Dummy	Yes	Yes	Yes	Yes
Region Time Trend	Yes	No	Yes	No
Year Dummy	No	Yes	No	Yes
R-squared	0.9854	0.9862	0.9549	0.9540
Number of observations	798	798	798	798

Note: Large regions refer to the 42 regions with the largest average population between 1998 and 2016. Small regions refer to the 42 regions with the smallest average population between 1998 and 2016. The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

difference between large and small regions in the regressions with year fixed effects and there are significant differences of the digitization coefficient for the regressions with time trends (Tables A42, A43).

Next I try to interpret the observed results. Considering the regression with time trends, the difference between large and small could be caused by differences in IT levels or IT level increases between larger and smaller regions. Considering the 30 smallest regions (on average below 20 000 inhabitants), the number of IT specialists has increased on average about 75% between 1998 and 2016. Across all regions the number has increased about 114% in this time period. In the 30 largest regions it has increased on average about 140%. Therefore, it is plausible to assume that the correlation coefficient is smaller in small regions, because the variation in digitization within regions over time is smaller than for the larger regions. Furthermore, the smaller region also have lower levels of IT specialist compared to their population size. For the 30 smallest regions, on average 0.68% of the population have IT education, for the 30 largest regions, on average 1.05 % of the population have IT education.

Considering the results of the functional form with year fixed effects, the correlation coefficient drops significantly in both subgroups. On the one hand, this could generally be an indication that there are relevant year unobserved effects. On the other hand, the decline of the coefficient is not the only interesting aspect of the results. For the large regions, the correlation coefficient is also smaller and less significant than in the same regression with the full sample of regions. For small regions the coefficient is slightly larger than in the full sample. My hypothesis is that this could be because large and small regions are implementing digital tools at a different speed. When digitization increases in all regions at the same rate in the same time intervals, then the digitization effect is likely to be controlled away by the year fixed effects. Consider an exaggerated example, if large regions all adopt a digital innovation at time t_1 then the effect is then taken up by the year fixed effect for year t_1. Controlling for year fixed effects requires that there is variation in the timing of the adoption of digitization between regions. The fact that the measured Digi coefficient is smaller in the sample with large regions, could therefore be an indication that the variation in digitization increase in each year is smaller in the pool of large regions than in the pool of small regions. This hypothesis is not rejected by a simple descriptive statistic. I compute the relative increase of IT specialists between 1998 and 2016 for each region and compare the averages of these numbers for the sample of large and the sample of small regions. The variation of this increase is larger in the sample of small regions than in large regions as measured by the standard deviation from the mean (Table A44 in Appendix 1).

Table 11: Weighted by population: FE Regression, Absolute Values

	Open Unemployed		Total Job Seekers	
	Trend	Year FE	Trend	Year FE
ln Digi	0.2939***	0.0579	0.2266***	0.0253
	(0.0796)	(0.0490)	(0.0541)	(0.0365)
ln Foreign born	-0.5213***	-0.0787	-0.2502	-0.1488**
	(0.1571)	(0.0900)	(0.1631)	(0.0722)
ln Old Unemployed	0.0236	0.0129	0.1613*	-0.1247
	(0.0313)	(0.0357)	(0.0932)	(0.0990)
ln Young Unemployed	0.0435	0.0027	0.2042***	-0.1252**
	(0.0356)	(0.0490)	(0.0596)	(0.0549)
ln Female Unemployed	-0.1413***	-0.2280***	-0.4624***	-0.3384***
	(0.0508)	(0.0597)	(0.1013)	(0.0877)
ln High Edu	-0.5282***	-0.4213*	-0.1769	-0.2783
	(0.1906)	(0.2318)	(0.1649)	(0.2094)
ln Low Edu	2.4204***	0.0762	1.6347***	-0.5469*
	(0.3470)	(0.3223)	(0.1831)	(0.2974)
ln Employment	-0.1439	-0.8963***	0.2270	-0.4150*
	(0.3362)	(0.2726)	(0.1850)	(0.2320)
ln Population	-0.5554	2.3377***	-1.1890***	1.6597***
	(0.4010)	(0.5642)	(0.4429)	(0.4941)
Year 2008	-0.1296***		-0.0662***	
	(0.0115)		(0.0109)	
Year 2009	-0.2253***		-0.1807***	
	(0.0190)		(0.0183)	
Time Trend	0.0618***		0.0327***	
	(0.0142)		(0.0117)	
ln Employed Job Seekers			0.0834	0.1724***
			(0.0508)	(0.0634)
ln Unemp. in Programs			-0.0900***	0.0965***
			(0.0241)	(0.0251)
Constant	-130.6589***	-11.4328***	-66.5133**	-1.9140
	(29.6508)	(2.6473)	(25.2089)	(2.2040)
Region Dummy	Yes	Yes	Yes	Yes
Region Time Trend	Yes	No	Yes	No
Year Dummy	No	Yes	No	Yes
R-squared	0.9936	0.9950	0.9657	0.9722
Number of observations	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open or total unemployed and seasonal fixed effects. Control variables are share values. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Assuming that there could be a difference between larger and smaller labor market regions concerning the potential digitization effect, it is also interesting to consider how digitization affects matching efficiency when the aggregate labor market is considered. While I employed unweighted regressions in all previous subsections, I now add a weighted regression, presented in Table 11.[16] This regression considers the

[16] Individual regression tables of each column and regressions with share values instead of absolute

full sample of regions, but observations are weighted according to population size. This measures the ceteris paribus correlation of digitization and matching efficiency in the aggregate Swedish labor market. When using time trends, the digitization coefficient now turns to be much larger than in an unweighted regression, with both open and total unemployed. This is an expected result given that the correlation was bigger in large regions. The weighted regression with year fixed effects exhibits a similar pattern as in the unweighted regression with year fixed effects. It shows a small an statistically insignificant coefficient. However, the coefficient is larger than in the unweighted regression (0.0579 vs. 0.0218) in the regression with open unemployed. This seems counter intuitive at first, since it implies that the correlation is stronger in large than in small regions. When the sample was split, larger regions exhibited a smaller digitization effect than small regions in the year fixed effects regression. But, this result strengthens the argument made earlier, that the digitization effect in the sample with only large regions is partly controlled away because of too little within-year between-region variation. Now in the full sample, there is more variation again. This implies that the digitization effect is stronger, not weaker in large regions.

Therefore, it can be concluded that on the aggregate Swedish labor market digitization is (ceteris paribus) positively correlated with matching efficiency. However, a large part of this correlation could be due to region-invariant unobserved events. Furthermore, the positive correlation between digitization and matching efficiency seems to be more pronounced in labor market regions with larger population sizes. This is probably caused by the, on average, larger increase in IT specialists in larger regions.

6.3 Summary and Discussion

When interpreting these results it is important to keep the methodological limitations in mind. First, I compute matching efficiency by taking the difference between the actual number of job hires (matches) in a month and the number predicted from unemployment and vacancy data. This measure is focused on the quantity of matches, not on the quality. Thereby, any estimated effect of digitization on matching efficiency does for example not entail whether the qualitative fit of candidates and employers has been increased.

Second, one of the main limitations of this study is that it is uncertain how well and how closely correlated the proxy variable is with the actual technology use in recruiting. The main identifying assumptions for using this proxy variable, is

values can be found in the appendix

that the number of IT specialists in each region is correlated with the use of digital tools in recruiting, and that this correlation is not influenced by any of the control variables in the final panel regression. The advantage of this proxy variable is that it is based on registry data and thus not prone to measurement error.

Third, my data ends in 2016 and the results do therefore not capture the most recent technological advancements in recruiting. Especially the use of AI based technologies only becomes widespread when my data sample ends. Therefore, analyses with more recent data might find more significant results.

Fourth, to interpret any of these correlations as the causal effect of recruiting, it needs to be assumed that recruiting is the only causal channel through which digitization affects job matching efficiency. Although I do not find it very probable, it is possible that there are other causal channels e.g. production productivity. Skill mismatch could be a causal channel too, but given that the estimated digitization coefficient is never significantly negative this seems unlikely. Further, it needs to be assumed that the causal channel is only working in one direction and there is no reversed causality. Moreover, there may be no unobserved events, which are correlated with digitization and matching efficiency. To minimize the risk of these validity threats, I robustness check my results with instrumenting the explanatory variables with lagged variables. These regressions point to a similar correlation coefficients as the results with non-lagged variables, which is a good indication for the internal validity of the results. Besides that, I have assessed different factors which are considered to influence matching efficiency in the literature and added corresponding controls and robustness checks.

Summarizing the empirical results, I estimated the correlation between digitization and matching efficiency within regions and found very mixed results. I do find some statistically significant positive correlation between digitization and matching efficiency when employing a fixed effects regression with time trends. The correlation coefficient decreases but maintains significance when other controls are added to the regression. When substituting the time trends with year fixed effects, the correlation coefficient is generally smaller and turns insignificant once more controls are added. Year fixed effects are able to account for unobserved events which are time-varying and region-invariant, while time trends are less granular at accounting for such unobserved effect. Therefore, the results with year fixed effects could indicate that the correlation is biased by unobserved effects. However, it can not be excluded that the results with only year fixed effects are confounded with regional economic trends such as structural economic shift. Another potential explanation is that there is very little within-year between-region correlation in the digitization variable, which makes the year fixed effect control for parts of the digitization effect.

I employ a regression with lagged explanatory variables as another method to account for potential unobserved events. I run regressions with both time trends and year fixed effects with lagged variables and they return remarkably similar results. In both cases we observe correlation coefficients which are statistically significant when few controls are included and gradually decrease and become statistically insignificant once all controls are added. Robustness checks with first difference regressions confirm a moderate magnitude of the correlation coefficient but are statistically insignificant too. When considering the different results for population-wise large and small regions, it can be inferred that the correlation is stronger in large regions than in small regions. This could be explained by the larger increase in the number of IT specialists in more populous regions. Overall, a significant effect of digitization on matching efficiency can not be shown, but not completely excluded either. If there was any effect, the results indicate that it is more likely to be a positive effect than a negative effect.

7 Conclusion

This study is motivated by the increased use of digital tools in recruiting and the relevance of this factor for the matching process on labor markets. My aim was therefore to study whether there is a measurable correlation between increases in firm digitization and job matching efficiency within Swedish labor market regions. While I do find indications that there exists some correlation between digitization and matching efficiency, the results are not robust to different functional forms of the regression. Especially unobserved region-invariant year-varying effects seem to decrease the explanatory power of the digitization coefficient. Nevertheless, the study can not exclude that there is any effect of digitization. Moreover, my measurement of digitization is only an approximation of the actual use of IT in recruiting. An analysis with a direct measurement of IT use in recruiting might therefore yield different results and could be a relevant point for further research.

Overall, my empirical results indicate that digitization does not have a very large and significant impact on job matching efficiency. Instead, compositional changes in the population, such as changes in education levels and the number of foreign born individuals seem to be much more predictive for job matching efficiency. In line with the previous literature, my results confirm that matching efficiency in Sweden has declined over time. If one is willing to assume that digitization has indeed increased the transparency and decreased the average hiring duration in the labor market, then these results imply that changes in the ease of job search might only had a minor influence on matching efficiency.

References

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, 128(6):2188–2244.

Alexandrakis, C. (2014). Sectoral differences in the use of information technology and matching efficiency in the us labour market. *Applied Economics*, 46(29):3562–3571.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Aranki, T. and Löf, M. (2008). Matchningsprocessen på den svenska arbetsmarknaden: En regional analys. *Penning-och valutapolitik*, 1(2008):48–58.

Arbetsförmedlingen (2022). Tidigare statistik, Arbetssökande 1996 - 2021. accessed on March, 30th via https://arbetsformedlingen.se/statistik/sok-statistik/tidigare-statistik.

Barnichon, R. and Figura, A. (2011). What drives matching efficiency? a tale of composition and dispersion. *FEDS Working Paper*.

Bhuller, M., Kostol, A., and Vigtel, T. (2019). How broadband internet affects labor market matching. *IZA Discussion Papers*, No. 12895.

Biagi, F. (2013). ICT and Productivity: A Review of the Literature. *Institute for Prospective Technological Studies Digital Economy Working Paper*.

Black, J. S. and van Esch, P. (2020). Ai-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, 63(2):215–226.

Bonthuis, B., Jarvis, V., and Vanhala, J. (2016). Shifts in euro area beveridge curves and their determinants. *IZA Journal of Labor Policy*, 5(1):1–17.

Börsch-Supan, A. H. (1991). Panel data analysis of the beveridge curve: is there a macroeconomic relation between the rate of unemployment and the vacancy rate? *Economica*, pages 279–297.

Bouvet, F. (2012). The beveridge curve in europe: new evidence using national and regional data. *Applied Economics*, 44(27):3585–3604.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Coles, M. and Petrongolo, B. (2008). A test between stock-flow matching and the random matching function approach. *International Economic Review*, 49(4):1113–1141.

Coles, M. G. and Smith, E. (1996). Cross-section estimation of the matching function: evidence from England and Wales. *Economica*, pages 589–597.

Czernich, N. (2011). Broadband infrastructure and unemployment-evidence for germany. accessed via https://doi.org/10.5282/ubm/epub.12279.

Daly, M. C., Hobijn, B., Şahin, A., and Valletta, R. G. (2012). A search and matching approach to labor markets: Did the natural rate of unemployment rise? *Journal of Economic Perspectives*, 26(3):3–26.

Eklund, J., Pettersson, L., and Karlsson, P. (2015). Arbetslöshet, vakanser och utbildning – Hur har matchningen på Svenska arbetsmarknad utvecklats sedan 1990-talskrisen? *Jönköpings International Business School Reasearch Reports*.

Forslund, A. and Johansson, K. (2007). Random and stock-flow models of labour market matching: Swedish evidence. *Working Paper*, 2007(11). Institute for Labor Market Policy Evaluation (IFAU).

Hall, R. E. and Schulhofer-Wohl, S. (2018). Measuring job-finding rates and matching efficiency with heterogeneous job-seekers. *American Economic Journal: Macroeconomics*, 10(1):1–32.

Higashi, Y. (2020). Effects of region-specific shocks on labor market tightness and matching efficiency: evidence from the 2011 tohoku earthquake in japan. *The Annals of Regional Science*, 65(1):193–219.

Hmoud, B., Laszlo, V., et al. (2019). Will artificial intelligence take over human resources recruitment and selection. *Network Intelligence Studies*, 7(13):21–30.

Horton, J. J. (2017). The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics*, 35(2):345–385.

Ignatova, M., Reilly, K., Spar, B., and Ilya, P. (2018). Global recruiting trends 2018: Four ideas changing how you hire. Accessed on 14.10.2021 via https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/resources/pdfs/chapter4-ai-global-recruiting-trends-2018.pdf.

Ilmakunnas, P., Maliranta, M., and Vainiomäki, J. (2005). Worker turnover and productivity growth. *Applied Economics Letters*, 12(7):395–398.

Jakobsen, V., Korpi, T., and Lorentzen, T. (2019). Immigration and integration policy and labour market attainment among immigrants to scandinavia. *European Journal of Population*, 35(2):305–328.

Järvenson, G. (2020). Matching efficiency in Sweden, 2000 to 2018. Accessed via https://www.diva-portal.org/smash/get/diva2:1446185/FULLTEXT01.pdf.

Karlström, A., Palme, M., and Svensson, I. (2008). The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in Sweden. *Journal of public economics*, 92(10-11):2071–2082.

Konjunkturinstitutet (2021). Lönebildningsrapporten 2021. Accessed on 28.03.2022 via https://www.konj.se/publikationer/lonebildningsrapporten/lonebildningsrapporten/2021-10-20-stor-utmaning-att-varaktigt-sanka-arbetslosheten.html.

Kroft, K. and Pope, D. G. (2014). Does online search crowd out traditional search and improve matching efficiency? evidence from craigslist. *Journal of Labor Economics*, 32(2):259–303.

Kuhn, P. and Mansour, H. (2014). Is internet job search still ineffective? *The Economic Journal*, 124(581):1213–1233.

Kuhn, P. and Skuterud, M. (2004). Internet job search and unemployment durations. *American Economic Review*, 94(1):218–232.

Ladu, M. G. (2012). The relationship between total factor productivity growth and employment: some evidence from a sample of European regions. *Empirica*, 39(4):513–524.

Layard, R., Nickell, S., and Jackman, R. (2005). *Unemployment: macroeconomic performance and the labour market*. Oxford University Press on Demand.

Lazear, E. P. and Spletzer, J. R. (2012). The united states labor market: Status quo or a new normal? *NBER Working Paper Series*, (18386). https://doi.org/10.3386/w18386.

Lisa, A. K. and Talla Simo, V. R. (2021). An in-depth study on the stages of ai in recruitment process of hrm and attitudes of recruiters and recruitees towards ai in sweden. accessed via https://www.diva-portal.org/smash/get/diva2:1566298/FULLTEXT01.pdf.

Mang, C. (2012). Online job search and matching quality. *Ifo Working Paper*, No. 147.

Mukoyama, T. (2014). The cyclicality of job-to-job transitions and its implications for aggregate productivity. *Journal of Economic Dynamics and Control*, 39:1–17.

Munoz, C., Smith, M., and Patil, D. (2016). *Big data: A report on algorithmic systems, opportunity, and civil rights*. Executive Office of the President. Accessed via https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

OECD (2009). Sickness, disability and work: Breaking the barriers sweden - will the recent reforms make it? *OECD Publishing Paris*. https://www.oecd.org/sweden/sicknessdisabilityandworkbreakingthebarrierssweden-willtherecentreformsmakeit.htm.

OECD (2016). Working together: Skills and labour market integration of immigrants and their children in sweden. *OECD Publishing Paris*. https://doi.org/10.1787/9789264257382-en.

Pedraza, P. (2008). Labour market matching efficiency in the czech republic transition. *William Davidson Institute Working Paper*.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature*, 39(2):390–431.

Shea, J. (1998). What do technology shocks do? *NBER macroeconomics annual*, 13:275–310.

Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics*, 15(3):370–399.

Spasova, S., Bouget, D., and Vanhercke, B. (2016). Sick pay and sickness benefit schemes in the European Union. *Background report for the Social Protection Committee's In-Depth Review on sickness benefits*.

Statistics Sweden (2022a). Enterprises and employees (FDB) by industrial classification SNI 2007 and size class. Year 2008 - 2013. accessed on March, 30th via https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__NV__NV0101/FDBR07/.

Statistics Sweden (2022b). Main provider of IT functions by accounting group. Share of enterprises. Year 2014 - 2017. accessed on March, 30th

via https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__NV__NV0116__NV0116H/ITSpecialistFunktion/.

Statistics Sweden (2022c). Reasons why enterprises have refrained from using AI technologies. Share of enterprises. Year 2021 . accessed on March, 30th via https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__NV__NV0116__NV0116M/AiTeknikerAvstatt/.

Statistics Sweden (2022d). Use of AI software or systems by way of acquiring them. Share of enterprises. Year 2021. accessed on March, 30th via https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__NV__NV0116__NV0116M/AiTeknikerForv/.

Tillväxtverket (2022). FA - Regioner. accessed on March, 30th via https://tillvaxtverket.se/statistik/regional-utveckling/regionala-indelningar/fa-regioner.html.

Wooldridge, J. (2018). *Introductory Econometrics - A modern Approach*. CENGAGE, 7th edition.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Yashiv, E. (2007). Labor search and matching in macroeconomics. *European Economic Review*, 51(8):1859–1895.

A Appendix 1

Table A12: Regression of Matching Functions 1996 - 2021

	Total Unemployed	Total Unemployed	Openly unemployed	Openly unemployed
$ln\ U_{it}^{in}$	0.2621***	0.0206	0.1559***	-0.0191
	(0.0225)	(0.0230)	(0.0204)	(0.0233)
$ln\ U_{it-1}^{stock}$	0.6099***	0.7451***	0.2695***	0.4008***
	(0.0402)	(0.0399)	(0.0366)	(0.0317)
$ln\ V_{it}^{in}$	0.0883***	0.0534***	0.1060***	0.0694***
	(0.0073)	(0.0065)	(0.0087)	(0.0080)
$ln\ V_{it-1}^{stock}$	0.0521***	0.0224***	0.0557***	0.0317***
	(0.0084)	(0.0048)	(0.0078)	(0.0056)
Constant	-1.5643***	-1.0567***	2.0682***	2.3260***
	(0.2722)	(0.3170)	(0.2345)	(0.2109)
Month Fixed Effects	-	Yes	-	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.9599	0.9780	0.9553	0.9757
Number of observations	22392	22392	22392	22392

Note: Standard errors are robust and clustered on region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A13: Summary Statistics Matching Efficiency

Regression Residuals ω	Obs	Mean	Std. dev.	Min	Max
Total Unemployed Seasonal	1,368	3.83e-10	.2718325	-1.094925	.9018487
Total Unemployed Year	1,368	-7.25e-11	.2363396	-.6442093	.9171241
Openly Unemployed Seasonal	1,368	8.46e-12	.8009249	-2.372055	2.335396
Openly Unemployed Year	1,368	-5.96e-10	.619674	-1.964862	1.747205

Table A14: FE Regression, IT Education and Professional Titles

	(1)
Education	2.6794***
	(0.0947)
Constant	-972.5999***
	(94.7380)
Region Fixed Effect	Yes
Year Fixed Effect	Yes
R-squared	0.9664
N	1152

Note: Based on numbers of Swedish FA regions. Standard errors are clustered by regions. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A15: Regression, IT Education and Professional Titles 2001 -2006

	2001	2002	2003	2004	2005	2006
Education	1.0361***	1.1367***	1.1257***	2.3859***	2.5332***	2.2625***
	(0.0164)	(0.0199)	(0.0199)	(0.0358)	(0.0365)	(0.0320)
Constant	-178.5138***	-212.4978***	-207.6984***	-389.0739***	-410.3694***	-451.2151***
	(39.8005)	(49.8829)	(51.2602)	(95.0234)	(100.6546)	(105.2614)
R-squared	0.9827	0.9790	0.9786	0.9845	0.9856	0.9862
N	72	72	72	72	72	72

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A16: Regression, IT Education and Professional Titles 2007 - 2012

	2007	2008	2009	2010	2011	2012
Education	2.3042***	2.3768***	2.4397***	2.4113***	2.3983***	2.3679***
	(0.0313)	(0.0309)	(0.0280)	(0.0279)	(0.0285)	(0.0269)
Constant	-465.6997***	-451.4874***	-441.0663***	-448.3136***	-457.0492***	-447.4887***
	(107.2727)	(109.5024)	(99.0584)	(101.7526)	(109.4399)	(107.3589)
R-squared	0.9872	0.9883	0.9909	0.9907	0.9902	0.9910
N	72	72	72	72	72	72

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A17: Regression, IT Education and Professional Titles 2013- 2016

	2013	2014	2015	2016
Education	2.3112***	1.7281***	1.7643***	2.1009***
	(0.0253)	(0.0210)	(0.0203)	(0.0243)
Constant	-424.0312***	-306.6677***	-310.7686***	-415.4382***
	(104.2784)	(89.7072)	(90.0504)	(112.0985)
R-squared	0.9917	0.9898	0.9908	0.9907
N	72	72	72	72

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A18: Summary Statistic, Share of Open Unemployed

	Mean	SD	Min	Max	N
Old Jobseekers	0.2624	0.0672	0.1015	0.5482	1368
Female Jobseekers	0.2326	0.0496	0.0818	0.4227	1368
Young Jobseekers	0.2961	0.0660	0.0877	0.5199	1368
All Jobseekers	0.2776	0.0486	0.1152	0.4624	1368

This table depicts the share of the total job seekers which is openly unemployed for each group.

Table A19: Summary Statistic Unemployment Pool Composition

		Mean	SD	Min	Max	N
Open Unemployed	Share Old	0.1748	0.0481	0.0481	0.4024	1368
	Share Young	0.1906	0.0356	0.0781	0.2979	1368
	Share Female	0.4139	0.0640	0.1521	0.5809	1368
Total Job Seekers	Share Old	0.1844	0.0363	0.0922	0.3294	1368
	Share Young	0.1796	0.0290	0.0987	0.2629	1368
	Share Female	0.4944	0.0530	0.3053	0.6599	1368
	Employed Job Seekers	0.2831	0.0911	0.0866	0.5791	1368
	Unemployed in Programs	0.1914	0.0661	0.0524	0.3912	1368

This table depicts the average share of the respective group in the respective unemployment pool, averaged over years and regions.

Table A20: Summary Statistic, Independent Variables as Shares

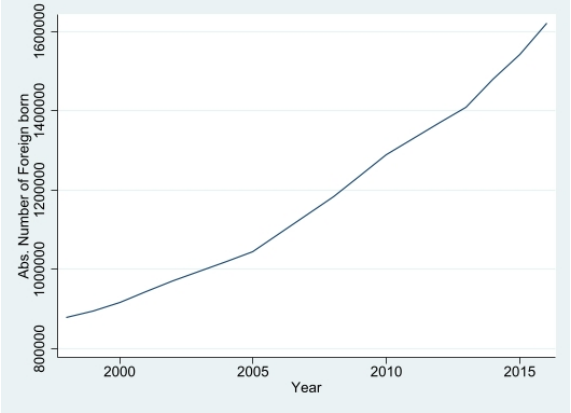
	Mean	SD	Min	Max	N
Digi	0.0086	0.0038	0.0000	0.0214	1368
Old Open Unemployed	0.0052	0.0023	0.0014	0.0242	1368
Old Total Unemployed	0.0199	0.0065	0.0073	0.0505	1368
Young Open Unemployed	0.0056	0.0017	0.0020	0.0136	1368
Young Total Unemployed	0.0195	0.0059	0.0052	0.0426	1368
Female Open Unemployed	0.0123	0.0039	0.0049	0.0350	1368
Female Total Unemployed	0.0535	0.0149	0.0241	0.1089	1368
Foreign Born	0.1098	0.0660	0.0243	0.4615	1368
High Education	0.1597	0.0501	0.0650	0.3589	1368
Medium Education	0.5021	0.0340	0.3954	0.6095	1368
Low Education	0.3147	0.0611	0.1530	0.4921	1368
Employed Job Seekers	0.0317	0.0153	0.0073	0.1041	1368
Unemployed in Program	0.0209	0.0093	0.0035	0.0617	1368

The variables are expressed as the absolute number of the respective group in year t in region i, divided by the population size of region i in year t.

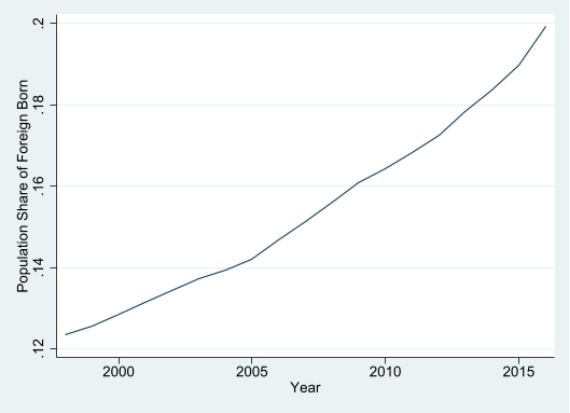
Table A21: Error Term Serial Correlation, FE Regression with Trends

	(1)
Residuals t-1	0.4419***
	(0.0261)
Constant	-0.0030
	(0.0023)
R-squared	0.1814
Number of observations	1296

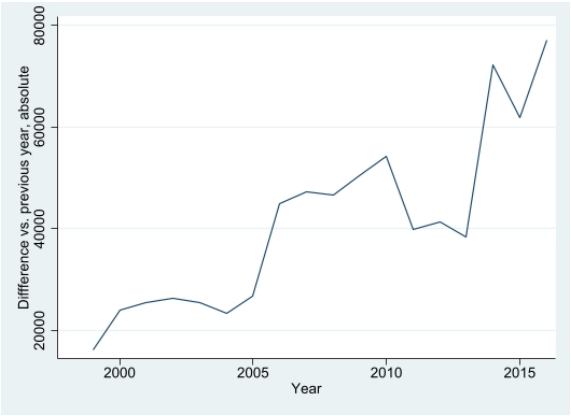
*** p < 0.01, ** p < 0.05, * p < 0.1.



(a) Absolute Number of Foreign Born

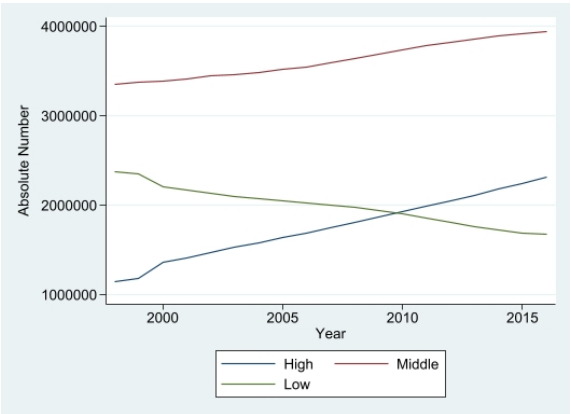


(b) Share of Foreign Born in Population

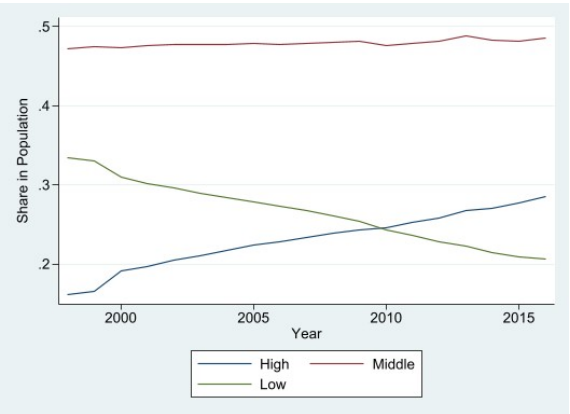


(c) Absolute Increase compared to Previous Year

Figure A3: Foreign Born Persons in Sweden over Time

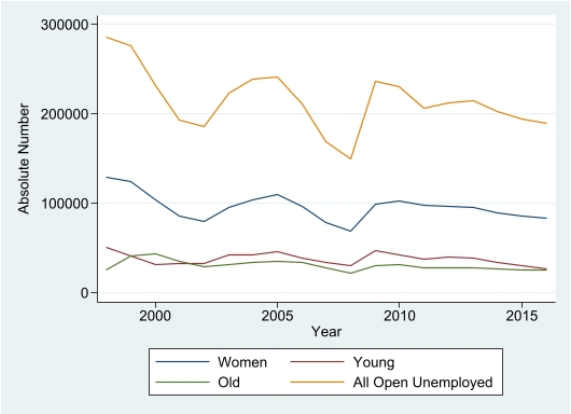


(a) Absolute Numbers



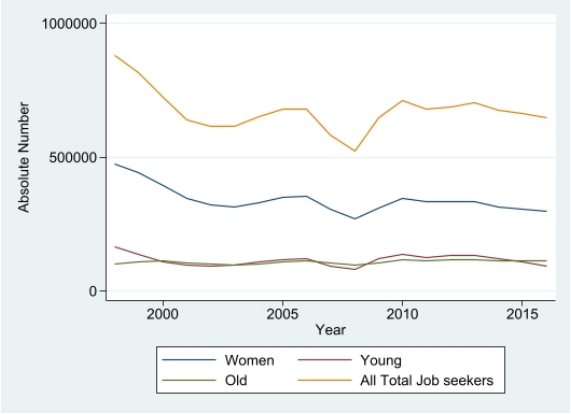
(b) Population Share

Figure A4: Education Levels in Sweden over Time

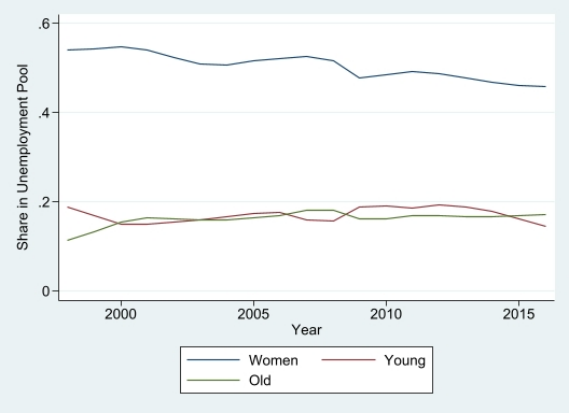


(a) Absolute Numbers, Openly Unemployed

(b) Shares in Pool of Openly Unemployed

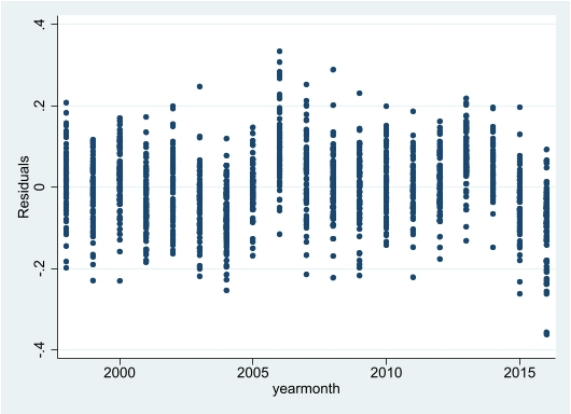


(c) Absolute Number, Total Job seekers

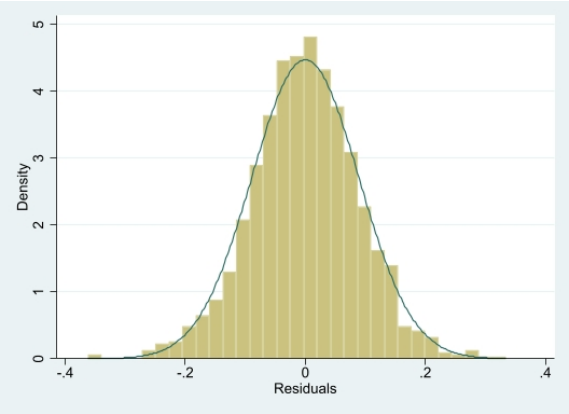


(d) Shares in Pool of Total Job seekers

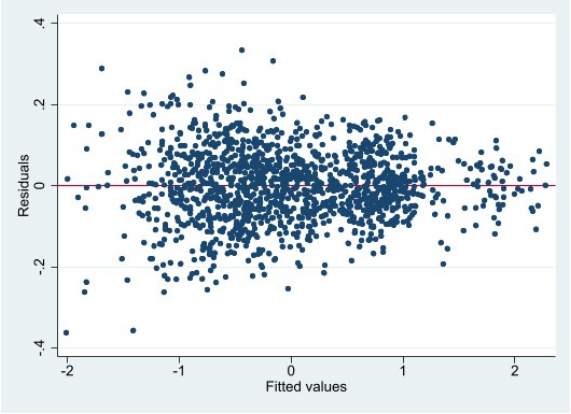
Figure A5: Composition of Unemployment Pool in Sweden over Time



(a) Predicted Errors over time



(b) Distribution of Predicted Errors



(c) Heteroskedasticity

Figure A6: Error Terms (FE Regression with Time Trend)

Table A22: FE Regression, Trends, Openly Unemployed, Shares

	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi	-0.4238***	0.2011***	0.1444***	0.1435***	0.0868***	0.0868**
	(0.0553)	(0.0357)	(0.0340)	(0.0318)	(0.0318)	(0.0342)
Year 2008		-0.0553***	-0.1065***	-0.1012***	-0.1149***	-0.1149***
		(0.0113)	(0.0112)	(0.0111)	(0.0104)	(0.0101)
Year 2009		-0.1933***	-0.2400***	-0.2365***	-0.2225***	-0.2225***
		(0.0115)	(0.0120)	(0.0119)	(0.0135)	(0.0128)
Time Trend		-0.0567***	0.0032	0.0238**	0.0035	0.0035
		(0.0011)	(0.0102)	(0.0118)	(0.0114)	(0.0148)
ln High Edu			0.3679**	0.1364	-0.1987	-0.1987
			(0.1624)	(0.1769)	(0.1763)	(0.1362)
ln Low Edu			2.2044***	1.8771***	1.5189***	1.5189***
			(0.2839)	(0.2938)	(0.2908)	(0.3057)
ln Foreign born				-0.4702***	-0.2558**	-0.2558**
				(0.1124)	(0.1018)	(0.1018)
ln Old Unemployed					0.1102***	0.1102***
					(0.0234)	(0.0189)
ln Young Unemployed					0.1670***	0.1670***
					(0.0314)	(0.0288)
ln Female Unemployed					0.1328**	0.1328**
					(0.0661)	(0.0559)
Constant	-3.3731***	113.6748***	-3.6652	-47.0722*	-6.3831	-6.3831
	(0.2935)	(2.4653)	(20.3195)	(23.8745)	(23.0566)	(29.6820)
Region Time Trend	No	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9410	0.9849	0.9864	0.9871	0.9880	0.9880
Number of observations	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with openly unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or the unemployment pool. Standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A23: FE Regression, Trends, Total Job seekers, Shares

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Digi	-0.2765***	0.2557***	0.1308***	0.1299***	0.0734***	0.0546*	0.0546**
	(0.0417)	(0.0400)	(0.0329)	(0.0310)	(0.0276)	(0.0280)	(0.0263)
Year2008		0.0069	-0.0554***	-0.0503***	-0.0634***	-0.0693***	-0.0693***
		(0.0100)	(0.0095)	(0.0098)	(0.0108)	(0.0100)	(0.0094)
Year2009		-0.1255***	-0.1813***	-0.1779***	-0.1734***	-0.1718***	-0.1718***
		(0.0112)	(0.0122)	(0.0121)	(0.0133)	(0.0129)	(0.0128)
Time Trend		-0.0273***	0.0214**	0.0411***	0.0251***	0.0258***	0.0258***
		(0.0013)	(0.0081)	(0.0099)	(0.0094)	(0.0090)	(0.0079)
ln High Edu			0.9922***	0.7705***	0.5096***	0.2351	0.2351
			(0.1413)	(0.1491)	(0.1775)	(0.1795)	(0.1741)
ln Low Edu			2.3645***	2.0509***	1.8757***	1.4174***	1.4174***
			(0.2273)	(0.2373)	(0.2477)	(0.2460)	(0.2178)
ln Foreign born				-0.4503***	-0.2737***	-0.2579***	-0.2579***
				(0.1061)	(0.0992)	(0.0956)	(0.0956)
ln Old Unemp.					0.2513***	0.2305***	0.2305***
					(0.0620)	(0.0613)	(0.0591)
ln Young Unemp.					0.1152***	0.1720***	0.1720***
					(0.0312)	(0.0381)	(0.0388)
ln Female Unemp.					0.1740	0.0278	0.0278
					(0.1104)	(0.1065)	(0.1016)
ln Emp. Job Seekers						0.1134**	0.1134***
						(0.0444)	(0.0388)
ln Unemp. in Programs						-0.0509**	-0.0509**
						(0.0232)	(0.0214)
Constant	-1.5269***	56.0180***	-37.6262**	-79.2025***	-46.9444**	-49.4751***	-49.4751***
	(0.2217)	(2.7557)	(16.1597)	(20.0105)	(18.9846)	(18.1702)	(16.0433)
Region Time Trend	No	Yes	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.6653	0.8833	0.9085	0.9138	0.9193	0.9226	0.9226
Number of obs.	1368	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or the unemployment pool. All standard errors are clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A24: FE Regression, Trends, Total Job Seekers, Abs.

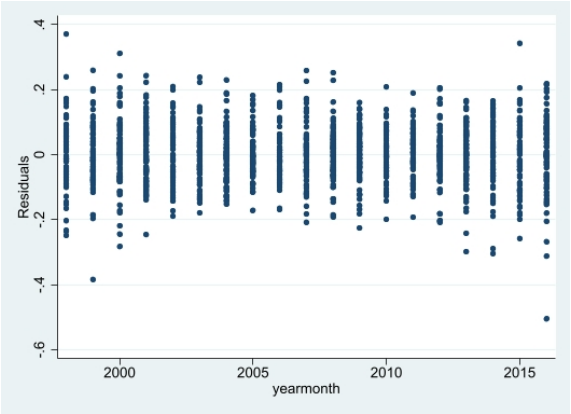
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Digi	0.2790***	0.1855***	0.1643***	0.1518***	0.1115***	0.0962***	0.0962***
	(0.0406)	(0.0378)	(0.0329)	(0.0327)	(0.0332)	(0.0307)	(0.0343)
Year2008	-0.0009	-0.0263***	-0.0360***	-0.0567***	-0.0666***	-0.0600***	-0.0600***
	(0.0103)	(0.0099)	(0.0094)	(0.0146)	(0.0126)	(0.0124)	(0.0106)
Year2009	-0.1264***	-0.1497***	-0.1612***	-0.1708***	-0.1733***	-0.1706***	-0.1706***
	(0.0111)	(0.0125)	(0.0118)	(0.0120)	(0.0120)	(0.0127)	(0.0110)
Time Trend	-0.0300***	-0.0138	0.0335***	0.0395***	0.0369***	0.0199	0.0199
	(0.0016)	(0.0104)	(0.0113)	(0.0117)	(0.0116)	(0.0136)	(0.0147)
ln High Edu		0.9022***	0.6497***	0.4019**	0.1471	0.0924	0.0924
		(0.1664)	(0.1594)	(0.1688)	(0.1675)	(0.1818)	(0.1583)
ln Low Edu		0.8915***	1.1998***	1.4916***	1.1384***	1.2005***	1.2005***
		(0.2778)	(0.2444)	(0.2521)	(0.2541)	(0.2854)	(0.2508)
ln Foreign born			-0.6500***	-0.5744***	-0.4654***	-0.3447***	-0.3447***
			(0.0897)	(0.0996)	(0.0946)	(0.1080)	(0.1124)
ln Old Unemp.				0.1303*	0.1366*	0.1403*	0.1403**
				(0.0743)	(0.0753)	(0.0761)	(0.0698)
ln Young Unemp.				0.0053	0.1021**	0.1111***	0.1111***
				(0.0297)	(0.0404)	(0.0393)	(0.0359)
ln Female Unemp.				-0.2162***	-0.3006***	-0.2925***	-0.2925***
				(0.0739)	(0.0795)	(0.0827)	(0.0724)
ln Emp. Job Seekers					0.1065**	0.0881**	0.0881**
					(0.0455)	(0.0416)	(0.0397)
ln Unemp. in Programs					-0.0807***	-0.0600**	-0.0600***
					(0.0235)	(0.0233)	(0.0225)
ln Population						-1.7043**	-1.7043**
						(0.7369)	(0.7331)
ln Employment						0.6155***	0.6155***
						(0.1809)	(0.1879)
Constant	59.5500***	15.8207	-76.2675***	-88.6614***	-80.1178***	-38.0303	-38.0303
	(3.0838)	(22.7149)	(23.6530)	(24.7911)	(24.6656)	(30.3148)	(32.2723)
Region Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.8874	0.8966	0.9109	0.9135	0.9192	0.9220	0.9220
Number of obs.	1368	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

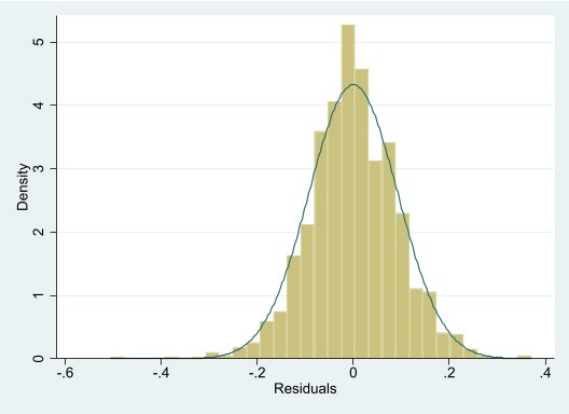
Table A25: Error Term Serial Correlation, Year FE

	(1)
Residuals t-1	0.6511***
	(0.0215)
Constant	-0.0000
	(0.0019)
R-squared	0.4139
Number of observations	1296

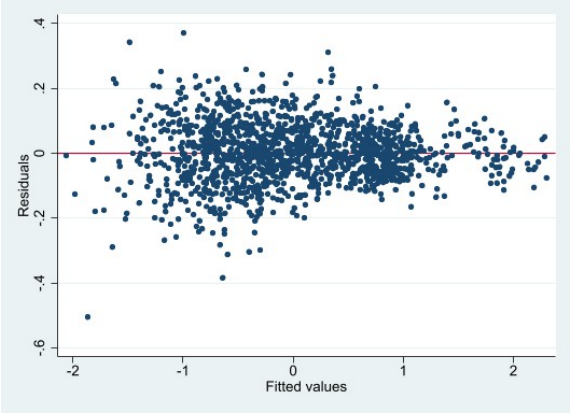
Note: *** p < 0.01, ** p < 0.05, * p < 0.1.



(a) Predicted Errors over time



(b) Distribution of Predicted Errors



(c) Heteroskedasticity

Figure A7: Error Terms (FE Regression with Year FE)

Table A26: FE Regression, Year FE and Time Trends, Open Unemployed, Abs.

	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi	-0.0296	-0.0193	-0.0218	-0.0166	-0.0235	-0.0235
	(0.0339)	(0.0314)	(0.0304)	(0.0305)	(0.0289)	(0.0315)
Time Trend	-0.0541***	-0.0285*	-0.0336**	-0.0261	-0.0273	-0.0273
	(0.0015)	(0.0155)	(0.0148)	(0.0160)	(0.0170)	(0.0179)
ln High Edu		-0.8231***	-0.6435***	-0.5772***	-0.6851***	-0.6851***
		(0.2007)	(0.2078)	(0.2018)	(0.2217)	(0.2177)
ln Low Edu		0.1490	0.2211	0.3245	0.1587	0.1587
		(0.3601)	(0.3481)	(0.3555)	(0.4742)	(0.4514)
ln Old Unemployed			0.0149	0.0121	0.0132	0.0132
			(0.0261)	(0.0258)	(0.0256)	(0.0214)
ln Young Unemployed			-0.0003	-0.0031	-0.0017	-0.0017
			(0.0304)	(0.0298)	(0.0298)	(0.0213)
ln Female Unemployed			-0.1634***	-0.1587***	-0.1579***	-0.1579***
			(0.0422)	(0.0422)	(0.0422)	(0.0393)
ln Foreign born				-0.1183	-0.1401	-0.1401
				(0.0918)	(0.1139)	(0.1016)
ln Employment					0.2034	0.2034
					(0.2419)	(0.2670)
ln Population					0.3558	0.3558
					(1.1777)	(1.0695)
Constant	107.3707***	59.8617*	68.9384**	53.5779	53.4701	53.4701
	(2.8759)	(32.9250)	(31.4675)	(33.7539)	(36.7592)	(36.6595)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Region Specific Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9909	0.9911	0.9917	0.9917	0.9917	0.9917
Number of observations	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A27: FE Regression, Year FE and Time Trends, Total Unemployed, Abs.

	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi	-0.0203	-0.0136	-0.0161	-0.0106	-0.0214	-0.0214
	(0.0324)	(0.0318)	(0.0312)	(0.0313)	(0.0301)	(0.0288)
Time Trend	-0.0198***	-0.0094	-0.0075	-0.0062	-0.0080	-0.0080
	(0.0014)	(0.0132)	(0.0131)	(0.0135)	(0.0154)	(0.0139)
ln High Edu		-0.6409***	-0.5235***	-0.4680***	-0.6551***	-0.6551***
		(0.1531)	(0.1482)	(0.1539)	(0.1703)	(0.1661)
ln Low Edu		-0.1343	0.1424	0.1513	-0.1429	-0.1429
		(0.3635)	(0.3439)	(0.3520)	(0.3240)	(0.3413)
ln Foreign born			-0.1225*	-0.1164	-0.1509	-0.1509
			(0.0729)	(0.0747)	(0.0997)	(0.0924)
ln Old Unemployed			-0.0045	-0.0277	-0.0164	-0.0164
			(0.0593)	(0.0616)	(0.0631)	(0.0676)
ln Young Unemployed			-0.0816*	-0.1085**	-0.0985**	-0.0985***
			(0.0409)	(0.0432)	(0.0420)	(0.0315)
ln Female Unemployed			-0.1069	-0.2116***	-0.2313***	-0.2313**
			(0.0684)	(0.0798)	(0.0851)	(0.0949)
ln Employed Job Seekers				0.0732*	0.0783**	0.0783**
				(0.0409)	(0.0386)	(0.0370)
ln Unemp. in Program				0.0622**	0.0714***	0.0714***
				(0.0277)	(0.0261)	(0.0234)
ln Employment					0.3535*	0.3535**
					(0.1873)	(0.1721)
ln Population					0.6272	0.6272
					(1.1741)	(1.1393)
Constant	39.5358***	23.2947	18.4066	15.6484	14.7780	14.7780
	(2.8288)	(29.0117)	(28.4049)	(29.4110)	(36.5036)	(32.9336)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Region Specific Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9409	0.9425	0.9455	0.9462	0.9468	0.9468
Number of observations	1368	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A28: FE Regression, Year Fixed Effects, Open Unemployed, Shares

	(1)	(2)	(3)	(4)	(5)
ln Digi	0.0893**	0.0762**	0.0713*	0.0303	0.0303
	(0.0376)	(0.0372)	(0.0379)	(0.0309)	(0.0337)
ln High Edu		-0.6296***	-0.6897***	-0.5530**	-0.5530**
		(0.2307)	(0.2486)	(0.2317)	(0.2455)
ln Low Edu		0.7926*	0.4591	-0.0089	-0.0089
		(0.4159)	(0.4657)	(0.3941)	(0.3585)
ln Old Unemployed			0.0138	-0.0026	-0.0026
			(0.0384)	(0.0372)	(0.0377)
ln Young Unemployed			0.0433	0.0277	0.0277
			(0.0473)	(0.0481)	(0.0489)
ln Female Unemployed			-0.2006***	-0.1556**	-0.1556**
			(0.0726)	(0.0759)	(0.0768)
ln Foreign born				-0.2381***	-0.2381***
				(0.0706)	(0.0760)
Constant	0.4329**	0.1168	-0.4945	-1.5625*	-1.5625*
	(0.1821)	(0.7929)	(0.9943)	(0.8149)	(0.8593)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9841	0.9848	0.9852	0.9859	0.9859
Number of observations	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or unemployment pool. All standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A29: FE Regression, Year Fixed Effects, Total Job seekers, Shares

	(1)	(2)	(3)	(4)	(5)
ln Digi	0.0574*	0.0502	0.0500	0.0137	0.0137
	(0.0324)	(0.0323)	(0.0314)	(0.0278)	(0.0342)
ln Unemp in Programs	0.0362	0.0345	0.0303	0.0202	0.0202
	(0.0315)	(0.0306)	(0.0318)	(0.0303)	(0.0305)
ln Employed Job Seekers	0.0974**	0.1433***	0.1383***	0.0858*	0.0858
	(0.0435)	(0.0446)	(0.0480)	(0.0460)	(0.0574)
ln Old Unemployed		0.0339	0.0181	-0.0068	-0.0068
		(0.0647)	(0.0659)	(0.0652)	(0.0766)
ln Young Unemployed		-0.0509	-0.0501	-0.0650	-0.0650
		(0.0494)	(0.0518)	(0.0532)	(0.0499)
ln Female Unemployed		-0.2561**	-0.2644**	-0.1775*	-0.1775
		(0.1184)	(0.1190)	(0.1019)	(0.1266)
ln High Edu			-0.3953*	-0.3065	-0.3065
			(0.2144)	(0.1895)	(0.1941)
ln Low Edu			-0.0069	-0.4985*	-0.4985*
			(0.3072)	(0.2853)	(0.2730)
ln Foreign born				-0.2006***	-0.2006***
				(0.0562)	(0.0573)
Constant	0.4692**	0.2799	-0.5182	-1.6714**	-1.6714**
	(0.1907)	(0.2516)	(0.7670)	(0.7163)	(0.7039)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9122	0.9138	0.9149	0.9190	0.9190
Number of observations	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or unemployment pool. All standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A30: FE Regression, Year Fixed Effects, Total Job seekers, Abs.

	(1)	(2)	(3)	(4)	(5)
ln Digi	0.0636**	0.0640**	0.0534**	0.0250	0.0250
	(0.0313)	(0.0287)	(0.0236)	(0.0221)	(0.0229)
ln Employed Job Seekers	0.0012	0.1712***	0.1804***	0.1220***	0.1220***
	(0.0282)	(0.0436)	(0.0432)	(0.0447)	(0.0422)
ln Unemp in Programs	-0.0374*	0.0808***	0.0677**	0.0768**	0.0768**
	(0.0206)	(0.0288)	(0.0332)	(0.0324)	(0.0308)
ln Old Unemployed		-0.0520	-0.0738	-0.0958	-0.0958
		(0.0733)	(0.0700)	(0.0682)	(0.0664)
ln Young Unemployed		-0.0766*	-0.0888**	-0.0931**	-0.0931**
		(0.0438)	(0.0436)	(0.0429)	(0.0450)
ln Female Unemployed		-0.3056***	-0.2998***	-0.2048**	-0.2048**
		(0.0899)	(0.0841)	(0.0818)	(0.0796)
ln High Edu			-0.3208*	-0.3587**	-0.3587**
			(0.1698)	(0.1771)	(0.1681)
ln Low Edu			0.4249**	-0.2146	-0.2146
			(0.1626)	(0.2654)	(0.3111)
ln Foreign born				-0.1946***	-0.1946***
				(0.0570)	(0.0563)
ln Employment				0.0427	0.0427
				(0.1732)	(0.1622)
ln Population				0.6933	0.6933
				(0.4789)	(0.5088)
Constant	-0.0936	1.0793***	0.1448	0.2407	0.2407
	(0.2902)	(0.3418)	(1.4979)	(1.8819)	(1.8627)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.9119	0.9178	0.9197	0.9233	0.9233
Number of observations	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A31: FD without Trends, Open Unemployed, Abs.

	(1)	(2)	(3)	(4)	(5)
Δln Digi	0.0146	0.1062*	0.1030*	0.0866*	0.0866*
	(0.0340)	(0.0579)	(0.0555)	(0.0506)	(0.0463)
ΔYear 2008	-0.1209***	-0.1235***	-0.1231***	-0.1538***	-0.1538***
	(0.0079)	(0.0081)	(0.0079)	(0.0080)	(0.0084)
ΔYear 2009	-0.2367***	-0.2361***	-0.2387***	-0.2314***	-0.2314***
	(0.0099)	(0.0099)	(0.0098)	(0.0100)	(0.0115)
Δln High Edu		-0.3508***	-0.1453	0.0826	0.0826
		(0.1172)	(0.1085)	(0.1317)	(0.1193)
Δln Low Edu		0.8433***	0.5551***	1.1918***	1.1918***
		(0.1178)	(0.1350)	(0.1765)	(0.1721)
Δln Foreign born			-0.4775***	-0.3667***	-0.3667***
			(0.0593)	(0.0608)	(0.0632)
Δln Old Unemployed				-0.0085	-0.0085
				(0.0137)	(0.0134)
Δln Young Unemployed				-0.0599***	-0.0599***
				(0.0184)	(0.0182)
Δln Female Unemployed				-0.0687**	-0.0687***
				(0.0262)	(0.0259)
Δln Employment				-0.1029	-0.1029
				(0.1348)	(0.1104)
Δln Population				-1.3581***	-1.3581***
				(0.3069)	(0.2534)
Standard Errors	Robust	Robust	Robust	Robust	Bootstrap
R-squared	0.3071	0.3906	0.4191	0.4530	0.4530
Number of observations	1296	1296	1296	1296	1296

Note: The independent variable is matching efficiency, computed as the residual of matching function open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A32: FD Regression, Total and Open Unemployed, Shares and Abs.

	Absolute Values		Population Shares			
	total	total	open	open	total	total
Δln Digi	0.0717*	0.0699	0.0887	0.0915*	0.0540	0.0531
	(0.0409)	(0.0425)	(0.0558)	(0.0538)	(0.0381)	(0.0393)
Δln Foreign born	-0.2479***	-0.2009**	-0.2836**	-0.3364***	-0.1772***	-0.1728*
	(0.0582)	(0.0971)	(0.1074)	(0.0700)	(0.0592)	(0.0981)
Δln Old Unemployed	0.0806*	0.0900*	0.0433**	0.0406**	0.1956***	0.2034***
	(0.0456)	(0.0486)	(0.0181)	(0.0169)	(0.0420)	(0.0476)
Δln Young Unemployed	0.0590*	0.0704**	0.0159	0.0127	0.1316***	0.1369***
	(0.0334)	(0.0337)	(0.0232)	(0.0222)	(0.0411)	(0.0449)
Δln Female Unemployed	-0.2364***	-0.2452***	0.0339	0.0367	-0.0119	-0.0188
	(0.0883)	(0.0900)	(0.0443)	(0.0418)	(0.1019)	(0.1031)
Δln High Edu	0.2583**	0.2634*	0.2633**	0.1917*	0.3316***	0.3556**
	(0.1263)	(0.1414)	(0.1215)	(0.1107)	(0.1251)	(0.1375)
Δln Low Edu	0.9680***	0.8481***	0.9481***	1.1730***	1.0347***	1.0222***
	(0.1843)	(0.2233)	(0.2284)	(0.1651)	(0.1851)	(0.2347)
Δln Employed Job Seekers	0.0823*	0.0727			0.1140***	0.1116**
	(0.0477)	(0.0518)			(0.0393)	(0.0423)
Δln Unemp in Programs	-0.0247	-0.0269			-0.0131	-0.0136
	(0.0194)	(0.0199)			(0.0167)	(0.0170)
Δln Population	-1.6568***	-1.9263***				
	(0.4182)	(0.5407)				
Δln Employment	0.0402	0.0660				
	(0.1299)	(0.1285)				
ΔYear2008	-0.0851***	-0.0848***	-0.1248***	-0.1261***	-0.0783***	-0.0780***
	(0.0096)	(0.0100)	(0.0082)	(0.0081)	(0.0082)	(0.0085)
ΔYear2009	-0.1901***	-0.1926***	-0.2391***	-0.2405***	-0.1740***	-0.1745***
	(0.0108)	(0.0115)	(0.0110)	(0.0104)	(0.0103)	(0.0109)
Constant		-0.0051	-0.0193**			0.0085
		(0.0094)	(0.0087)			(0.0080)
Region Dummy	No	Yes	Yes	No	No	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust	Robust
R-squared	0.3529	0.3229	0.3829	0.4279	0.3599	0.3294
Number of observations	1296	1296	1296	1296	1296	1296

Note: The independent variable is matching efficiency, computed as the residual of matching function of total or open unemployed and seasonal fixed effects. Control variables are either absolute values or expressed as shares of the population or unemployment pool. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A33: Autoregression Error Term, Open Unemp, Year FE

	(1)	(2)	(3)	(4)
Residuals t-1	0.6511***			
	(0.0215)			
Residuals t-2		0.4156***		
		(0.0266)		
Residuals t-3			0.2074***	
			(0.0293)	
Residuals t-4				0.0145
				(0.0308)
Constant	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0019)	(0.0024)	(0.0026)	(0.0027)
R-squared	0.4139	0.1663	0.0417	0.0002
Number of observations	1296	1224	1152	1080

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A34: Autoregression Error Term, Total Unemp, Year FE

	(1)	(2)	(3)	(4)
Residuals t-1	0.5464***			
	(0.0238)			
Residuals t-2		0.3085***		
		(0.0277)		
Residuals t-3			0.1195***	
			(0.0297)	
Residuals t-4				-0.0292
				(0.0305)
Constant	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0018)	(0.0020)	(0.0022)	(0.0022)
R-squared	0.2887	0.0920	0.0139	0.0008
Number of observations	1296	1224	1152	1080

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A35: Autoregression Error Term, Open Unemp, Trend

	(1)	(2)
Residuals t-1	0.4419***	
	(0.0261)	
Residuals t-2		0.0004
		(0.0300)
Constant	-0.0030	-0.0002
	(0.0023)	(0.0026)
R-squared	0.1814	0.0000
Number of observations	1296	1224

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A36: Autoregression Error Term, Total Unemp, Trend

	(1)	(2)
Residuals t-1	0.3903***	
	(0.0267)	
Residuals t-2		-0.0022
		(0.0298)
Constant	-0.0017	0.0004
	(0.0020)	(0.0022)
R-squared	0.1420	0.0000
Number of observations	1296	1224

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A37: Autoregression Digi

	(1)	(2)
ln Digi t-2	1.0042***	
	(0.0027)	
ln Digi t-4		1.0129***
		(0.0037)
Constant	0.0540***	0.0732***
	(0.0146)	(0.0201)
R-squared	0.9912	0.9855
Number of observations	1224	1080

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A38: Lagged: FE Regression, Year FE, Total Job Seekers, Abs.

	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi t-4	0.0526	0.0561	0.0420	0.0278	0.0153	0.0185
	(0.0343)	(0.0337)	(0.0286)	(0.0269)	(0.0277)	(0.0267)
ln Old Unemp t-4		-0.0766	-0.1136*	-0.1110*	-0.1075	-0.1150*
		(0.0758)	(0.0654)	(0.0654)	(0.0666)	(0.0629)
ln Young Unemp t-4		0.0575	0.0272	0.0112	0.0314	0.0181
		(0.0644)	(0.0568)	(0.0564)	(0.0540)	(0.0548)
ln Female Unemp t-4		0.1251	0.1990**	0.2025**	0.1662**	0.0513
		(0.0987)	(0.0852)	(0.0854)	(0.0823)	(0.1144)
ln Foreign born t-4			-0.2311***	-0.2308***	-0.2719***	-0.2475***
			(0.0512)	(0.0649)	(0.0622)	(0.0579)
ln High Edu t-4				0.1144	-0.1740	-0.1879
				(0.1729)	(0.1732)	(0.1690)
ln Low Edu t-4				0.0889	-0.8315*	-0.7867*
				(0.2304)	(0.4293)	(0.4115)
ln Employment t-4					-0.0168	-0.0541
					(0.1893)	(0.1772)
ln Population t-4					1.5355***	1.5639***
					(0.5432)	(0.5141)
ln Emp. Job Seekers t-4						0.0916*
						(0.0520)
ln Unemp in Program t-4						0.0213
						(0.0303)
Constant	-0.2964*	-1.1249**	0.6618	-0.9985	-5.2997***	-5.5125***
	(0.1738)	(0.4632)	(0.5450)	(1.6347)	(1.9013)	(2.0027)
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.9216	0.9243	0.9287	0.9292	0.9310	0.9316
Number of observations	1080	1080	1080	1080	1080	1080

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A39: Lagged Values: FE Regression, Trends, Openly Unemployed, Abs.

	(1)	(2)	(3)	(4)	(5)
ln Digi t-2	0.1019**	0.1004**	0.1048**	0.0688*	0.0048
	(0.0393)	(0.0391)	(0.0403)	(0.0383)	(0.0395)
Year 2008	-0.0583***	-0.0654***	-0.0787***	-0.0948***	-0.1092***
	(0.0123)	(0.0123)	(0.0121)	(0.0120)	(0.0096)
Year 2009	-0.1989***	-0.1908***	-0.1974***	-0.2179***	-0.2357***
	(0.0119)	(0.0118)	(0.0117)	(0.0127)	(0.0129)
Time Trend	-0.0609***	-0.0630***	-0.0371***	-0.0484***	-0.0277
	(0.0015)	(0.0017)	(0.0047)	(0.0089)	(0.0168)
ln Old Unemp t-2		0.0513**	0.0112	0.0262	0.0026
		(0.0210)	(0.0234)	(0.0247)	(0.0259)
ln Young Unemp t-2		0.1932***	0.1495***	0.1474***	0.0755**
		(0.0251)	(0.0271)	(0.0306)	(0.0318)
ln Female Unemp t-2		-0.1037***	-0.0352	-0.0435	0.0377
		(0.0295)	(0.0378)	(0.0479)	(0.0462)
ln Foreign born t-2			-0.5677***	-0.5039***	-0.3894***
			(0.0926)	(0.1148)	(0.1202)
ln Population t-2				-0.4624	-1.5775**
				(0.7635)	(0.7098)
ln Employment t-2				0.8853***	0.4182*
				(0.2264)	(0.2388)
ln High Edu t-2					0.9640***
					(0.2131)
ln Low Edu t-2					1.3792***
					(0.3859)
Constant	120.9731***	124.7776***	75.7678***	95.1549***	50.3911
	(2.9654)	(3.3434)	(8.9505)	(21.9087)	(36.2108)
Region Time Trend	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes
R-squared	0.9844	0.9857	0.9866	0.9871	0.9881
Number of observations	1224	1224	1224	1224	1224

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A40: Lagged Values: FE Regression, Trends, Total Job Seekers, Abs.

	(1)	(2)	(3)	(4)	(5)	(6)
ln Digi t-2	0.1200***	0.0522	0.0704*	0.0395	0.0253	-0.0079
	(0.0405)	(0.0408)	(0.0417)	(0.0409)	(0.0367)	(0.0356)
Year 2008	0.0008	0.0083	-0.0073	-0.0289**	-0.0492***	-0.0593***
	(0.0108)	(0.0116)	(0.0118)	(0.0128)	(0.0124)	(0.0112)
Year 2009	-0.1337***	-0.1304***	-0.1385***	-0.1584***	-0.1769***	-0.1701***
	(0.0111)	(0.0117)	(0.0111)	(0.0113)	(0.0151)	(0.0155)
Time Trend	-0.0289***	-0.0391***	-0.0097	-0.0235**	-0.0037	0.0004
	(0.0016)	(0.0027)	(0.0063)	(0.0100)	(0.0115)	(0.0144)
ln Old Unemp t-2		0.2811***	0.1548*	0.1967**	0.1037	-0.0101
		(0.0783)	(0.0809)	(0.0803)	(0.0770)	(0.0694)
ln Young Unemp t-2		0.0886**	0.0664*	0.0787*	0.0906*	0.0354
		(0.0416)	(0.0388)	(0.0438)	(0.0455)	(0.0471)
ln Female Unemp t-2		-0.4436***	-0.2950***	-0.2962***	-0.5078***	-0.2536***
		(0.0818)	(0.0881)	(0.0938)	(0.0921)	(0.0941)
ln Foreign born t-2			-0.5687***	-0.4836***	-0.4201***	-0.3538***
			(0.1027)	(0.1274)	(0.1182)	(0.1233)
ln Population t-2				-0.8018	-0.0273	-1.4887**
				(0.7092)	(0.7446)	(0.7407)
ln Employment t-2				0.9301***	0.8723***	0.5631***
				(0.1934)	(0.1538)	(0.1570)
ln Emp Job Seekers t-2					0.2729***	0.1929***
					(0.0593)	(0.0568)
ln Unemp in Program t-2					0.0259	0.0543**
					(0.0231)	(0.0233)
ln Low Edu t-2						1.0180***
						(0.2883)
ln High Edu t-2						1.0314***
						(0.2184)
Constant	57.7244***	78.9599***	22.7893*	49.1203**	3.4644	-4.3885
	(3.0652)	(5.5073)	(12.2533)	(23.0847)	(26.9857)	(31.9473)
Region Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.8850	0.8951	0.9021	0.9072	0.9132	0.9191
Number of observations	1224	1224	1224	1224	1224	1224

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A41: Large and Small Regions, Total Job seekers, Abs.

	42 Largest Regions		42 Smallest Regions	
	Trend	Year FE	Trend	Year FE
ln Digi	0.1517***	-0.0363	0.0714**	0.0375
	(0.0460)	(0.0445)	(0.0307)	(0.0227)
ln Foreign born	-0.6293***	-0.1858*	-0.3459***	-0.1866***
	(0.1433)	(0.0985)	(0.1231)	(0.0635)
ln Old Unemployed	0.0607	-0.1841	0.1188	-0.0948*
	(0.0907)	(0.1659)	(0.0874)	(0.0548)
ln Young Unemployed	0.1984***	-0.0847	0.0694	-0.0829*
	(0.0494)	(0.0682)	(0.0472)	(0.0442)
ln Female Unemployed	-0.3558***	-0.1694	-0.2444**	-0.2432**
	(0.0922)	(0.1146)	(0.0956)	(0.1052)
ln High Edu	-0.2635	-0.8353***	0.4999*	-0.2374
	(0.1654)	(0.2948)	(0.2917)	(0.2380)
ln Low Edu	1.2330***	-0.5532	0.9762***	0.0646
	(0.3236)	(0.4215)	(0.3510)	(0.3352)
Year 2008	-0.0700***		-0.0653***	
	(0.0153)		(0.0169)	
Year 2009	-0.1971***		-0.1531***	
	(0.0158)		(0.0184)	
ln Employed Job Seekers	0.0897	0.1059	0.0815*	0.1181**
	(0.0617)	(0.0671)	(0.0467)	(0.0509)
ln Unemp in Programs	-0.0710***	0.0673	-0.0581*	0.0632
	(0.0258)	(0.0408)	(0.0306)	(0.0412)
Time Trend	0.0248**		0.0024	
	(0.0106)		(0.0155)	
ln Employment	0.5781**	-0.3910	0.6188***	0.1047
	(0.2536)	(0.3003)	(0.1949)	(0.2033)
ln Population	-1.5503*	2.1743***	-1.7719*	-0.1798
	(0.7973)	(0.7375)	(0.9253)	(0.7214)
Constant	-43.3350*	-2.6568	-3.2490	4.4524
	(23.9075)	(3.4431)	(34.7335)	(3.5216)
Region Dummy	Yes	Yes	Yes	Yes
Region Time Trend	Yes	No	Yes	No
Year Dummy	No	Yes	No	Yes
R-squared	0.9291	0.9317	0.8626	0.8667
Number of observations	798	798	798	798

Note: Large regions refer to the 42 regions with the largest average population between 1998 and 2016. Small regions refer to the 42 regions with the smallest average population between 1998 and 2016. The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A42: Large Regions, FE Regression, Shares

	Time Trend		Year FE	
	Open	Total	Open	Total
ln Digi	0.1175**	0.0868*	-0.0448	-0.0198
	(0.0503)	(0.0434)	(0.0596)	(0.0429)
ln Foreign born	-0.4474***	-0.2603*	-0.2905**	-0.1620
	(0.1303)	(0.1516)	(0.1099)	(0.1053)
ln Old Unemployed	0.0992***	0.2657***	-0.0032	0.0180
	(0.0247)	(0.0674)	(0.0727)	(0.1570)
ln Young Unemployed	0.1669***	0.2925***	0.0905	0.0612
	(0.0409)	(0.0455)	(0.0639)	(0.0695)
ln Female Unemployed	0.3281***	0.2017**	0.0804	0.0058
	(0.0638)	(0.0997)	(0.1290)	(0.1882)
ln High Edu	-0.6958***	-0.1505	-0.7748**	-0.7830**
	(0.1418)	(0.1555)	(0.3566)	(0.3167)
ln Low Edu	1.1982***	1.0498***	0.2196	-0.7422*
	(0.2741)	(0.3284)	(0.4639)	(0.4129)
Year 2008	-0.1250***	-0.0701***		
	(0.0107)	(0.0125)		
Year 2009	-0.2029***	-0.1721***		
	(0.0133)	(0.0144)		
Time Trend	0.0201**	0.0205**		
	(0.0078)	(0.0092)		
ln Employed Job Seekers		0.1377**		0.0650
		(0.0545)		(0.0670)
ln Unemp in Program		-0.0569**		0.0442
		(0.0211)		(0.0361)
Constant	-40.2773**	-39.2919**	-1.2240	-2.2940*
	(15.8735)	(18.4228)	(0.9070)	(1.1833)
Region Dummy	Yes	Yes	Yes	Yes
Region Time Trend	Yes	Yes	No	No
Year Dummy	No	No	Yes	Yes
R-squared	0.9861	0.9325	0.9844	0.9259
Number of observations	798	798	798	798

Note: Large regions refer to the 42 regions with the largest average population between 1998 and 2016. The independent variable is matching efficiency, computed as the residual of matching function with open or total unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or the unemployment pool. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A43: Small Regions, FE Regression, Shares

	Time Trend		Year FE	
	Open	Total	Open	Total
ln Digi	0.0596*	0.0309	0.0171	0.0097
	(0.0321)	(0.0284)	(0.0324)	(0.0316)
ln Foreign born	-0.2196*	-0.2577**	-0.2039**	-0.2248***
	(0.1244)	(0.1036)	(0.0848)	(0.0603)
ln Old Unemployed	0.1055***	0.2106***	0.0339	-0.0015
	(0.0318)	(0.0747)	(0.0429)	(0.0631)
ln Young Unemployed	0.1558***	0.1441***	0.0549	-0.0673
	(0.0381)	(0.0477)	(0.0566)	(0.0616)
ln Female Unemployed	0.1240*	0.0621	-0.1605*	-0.1302
	(0.0725)	(0.1199)	(0.0806)	(0.1266)
ln High Edu	0.5098*	0.7328**	-0.3472	-0.1583
	(0.2992)	(0.2906)	(0.3207)	(0.2533)
ln Low Edu	1.2715***	1.2398***	-0.1722	-0.3975
	(0.3886)	(0.3007)	(0.5660)	(0.3817)
Year 2008	-0.1120***	-0.0727***		
	(0.0156)	(0.0158)		
Year 2009	-0.2170***	-0.1591***		
	(0.0196)	(0.0189)		
Time Trend	-0.0256*	0.0070		
	(0.0150)	(0.0112)		
ln Employed Job Seekers		0.1045**		0.0729
		(0.0503)		(0.0526)
ln Unemp in Program		-0.0518		-0.0018
		(0.0313)		(0.0376)
Constant	53.0206*	-10.9417	-1.8549	-1.5172
	(30.3614)	(22.6150)	(1.1444)	(0.9925)
Region Dummy	Yes	Yes	Yes	Yes
Region Time Trend	Yes	Yes	No	No
Year Dummy	No	No	Yes	Yes
R-squared	0.9564	0.8633	0.9507	0.8560
Number of observations	798	798	798	798

Note: Small regions refer to the 42 regions with the smallest average population between 1998 and 2016. The independent variable is matching efficiency, computed as the residual of matching function with open or total unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or the unemployment pool. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A44: Small and Large Regions, Increase in IT specialists 1998-2016

	Obs	Mean	Std. dev.	Min	Max
Small Regions	42	0.9539034	0.9792185	-0.75	3.454545
Large Regions	42	1.418888	0.7569524	0.3525641	3.494915

The increase in IT specialists is computed as the relative increase of IT specialists in the year 2016, compared with the year 1998 for each labor market region.

Table A45: Weighted FE Regression, Time Trend, Open Unemployed, Abs.

	(1)	(2)	(3)	(4)
ln Digi	0.2485***	0.3624***	0.3152***	0.2939***
	(0.0500)	(0.0795)	(0.0708)	(0.0796)
Year 2008	-0.0522***	-0.0962***	-0.1083***	-0.1296***
	(0.0108)	(0.0100)	(0.0112)	(0.0115)
Year 2009	-0.1684***	-0.2107***	-0.2252***	-0.2253***
	(0.0179)	(0.0178)	(0.0196)	(0.0190)
Time Trend	-0.0597***	0.0153	0.0253***	0.0618***
	(0.0020)	(0.0099)	(0.0084)	(0.0142)
ln High Edu		-0.3093*	-0.5362***	-0.5282***
		(0.1697)	(0.1405)	(0.1906)
ln Low Edu		1.9224***	2.0236***	2.4204***
		(0.2889)	(0.2448)	(0.3470)
ln Old Unemployed			0.0749**	0.0236
			(0.0319)	(0.0313)
ln Young Unemployed			0.1036***	0.0435
			(0.0323)	(0.0356)
ln Female Unemployed			-0.2133***	-0.1413***
			(0.0447)	(0.0508)
ln Foreign born				-0.5213***
				(0.1571)
ln Employment				-0.1439
				(0.3362)
ln Population				-0.5554
				(0.4010)
Constant	118.0483***	-44.0104*	-63.0760***	-130.6589***
	(3.8860)	(22.2617)	(18.8331)	(29.6508)
Region Specific Time Trend	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust
R-squared	0.9913	0.9929	0.9932	0.9936
Number of observations	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A46: Weighted FE Regression, Year FE, Open Unemployed, Abs.

	(1)	(2)	(3)	(4)
ln Digi	0.1160*	0.0592	0.0499	0.0579
	(0.0595)	(0.0531)	(0.0564)	(0.0490)
ln High Edu		0.0626	-0.0515	-0.4213*
		(0.2419)	(0.2081)	(0.2318)
ln Low Edu		0.4612**	1.0444***	0.0762
		(0.1985)	(0.1965)	(0.3223)
ln Old Unemployed			0.0328	0.0129
			(0.0345)	(0.0357)
ln Young Unemployed			-0.0035	0.0027
			(0.0477)	(0.0490)
ln Female Unemployed			-0.2455***	-0.2280***
			(0.0591)	(0.0597)
ln Foreign born				-0.0787
				(0.0900)
ln Employment				-0.8963***
				(0.2726)
ln Population				2.3377***
				(0.5642)
Constant	0.3015	-5.1996***	-8.6426***	-11.4328***
	(0.4719)	(1.8078)	(1.5652)	(2.6473)
Year Dummy	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust
R-squared	0.9936	0.9940	0.9946	0.9950
Number of observations	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A47: Weighted FE Regression, Time Trend, Total Job seekers, Abs.

	(1)	(2)	(3)	(4)	(5)
ln Digi	0.4779***	0.3406***	0.2970***	0.2717***	0.2266***
	(0.0659)	(0.0573)	(0.0571)	(0.0586)	(0.0541)
Year 2008	0.0042	-0.0250	-0.0584***	-0.0689***	-0.0662***
	(0.0121)	(0.0154)	(0.0092)	(0.0112)	(0.0109)
Year 2009	-0.1110***	-0.1466***	-0.1785***	-0.1853***	-0.1807***
	(0.0179)	(0.0209)	(0.0168)	(0.0180)	(0.0183)
Time Trend	-0.0376***	0.0115	0.0214**	0.0326**	0.0327***
	(0.0026)	(0.0131)	(0.0093)	(0.0124)	(0.0117)
ln High Edu		0.6059***	-0.0080	0.0644	-0.1769
		(0.1616)	(0.1906)	(0.1664)	(0.1649)
ln Low Edu		1.5432***	1.8726***	2.2889***	1.6347***
		(0.3207)	(0.2564)	(0.2907)	(0.1831)
ln Old Unemployed			0.3346***	0.1457	0.1613*
			(0.1223)	(0.0920)	(0.0932)
ln Young Unemployed			0.1254***	0.1247***	0.2042***
			(0.0273)	(0.0436)	(0.0596)
ln Female Unemployed			-0.5949***	-0.3978***	-0.4624***
			(0.1291)	(0.0997)	(0.1013)
ln Foreign born				-0.2714	-0.2502
				(0.1789)	(0.1631)
ln Employment				0.4962**	0.2270
				(0.2084)	(0.1850)
ln Population				-2.8397***	-1.1890***
				(0.4590)	(0.4429)
ln Employed Job Seekers					0.0834
					(0.0508)
ln Unemp in Program					-0.0900***
					(0.0241)
Constant	74.2258***	-37.9245	-55.3141***	-61.2189**	-66.5133**
	(5.1293)	(28.1768)	(20.7483)	(25.9460)	(25.2089)
Region Time Trend	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust
R-squared	0.9466	0.9517	0.9568	0.9631	0.9657
Number of observations	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A48: Weighted FE Regression, Year FE, Total Job seekers, Abs.

	(1)	(2)	(3)	(4)	(5)
ln Digi	0.0474	0.0433	0.0427	0.0422	0.0253
	(0.0431)	(0.0367)	(0.0415)	(0.0367)	(0.0365)
ln High Edu		0.3855	0.1444	-0.2915	-0.2783
		(0.2537)	(0.2035)	(0.2090)	(0.2094)
ln Low Edu		-0.5997***	0.1853	-0.8142***	-0.5469*
		(0.2111)	(0.1867)	(0.2762)	(0.2974)
ln Old Unemployed			-0.0010	-0.0690	-0.1247
			(0.0914)	(0.0972)	(0.0990)
ln Young Unemployed			-0.0495	-0.0387	-0.1252**
			(0.0486)	(0.0502)	(0.0549)
ln Female Unemployed			-0.2159**	-0.1744**	-0.3384***
			(0.0956)	(0.0800)	(0.0877)
ln Foreign born				-0.1850**	-0.1488**
				(0.0708)	(0.0722)
ln Employment				-0.4446*	-0.4150*
				(0.2490)	(0.2320)
ln Population				2.0258***	1.6597***
				(0.4820)	(0.4941)
ln Employed Job Seekers					0.1724***
					(0.0634)
ln Unemp in Program					0.0965***
					(0.0251)
Constant	-0.0440	2.4825	-1.2380	-3.1918	-1.9140
	(0.3414)	(1.6133)	(1.4502)	(2.3517)	(2.2040)
Year Dummy	Yes	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust	Robust
R-squared	0.9623	0.9652	0.9691	0.9711	0.9722
Number of observations	1368	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with total unemployed and seasonal fixed effects. Control variables are absolute values. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A49: Weighted FE Regression, Time Trend, Open Unemployed, Shares

	(1)	(2)	(3)	(4)
ln Digi	0.2610***	0.3677***	0.2521***	0.2528***
	(0.0583)	(0.0884)	(0.0697)	(0.0701)
Year 2008	-0.0512***	-0.1090***	-0.1077***	-0.1067***
	(0.0106)	(0.0097)	(0.0132)	(0.0124)
Year 2009	-0.1718***	-0.2273***	-0.1875***	-0.1873***
	(0.0168)	(0.0176)	(0.0184)	(0.0187)
Time Trend	-0.0582***	0.0232**	-0.0044	0.0012
	(0.0019)	(0.0102)	(0.0062)	(0.0130)
ln High Edu		-0.4005**	-0.6689***	-0.6756***
		(0.1618)	(0.1059)	(0.1108)
ln Low Edu		2.3282***	1.5594***	1.5488***
		(0.3631)	(0.2146)	(0.2192)
ln Old Unemployed			0.1245***	0.1191***
			(0.0309)	(0.0329)
ln Young Unemployed			0.1788***	0.1707***
			(0.0420)	(0.0489)
ln Female Unemployed			0.3423***	0.3343***
			(0.0721)	(0.0763)
ln Foreign born				-0.1017
				(0.1918)
Constant	117.0974***	-44.0576**	10.3385	-1.1203
	(4.0942)	(20.2101)	(12.3641)	(26.6540)
Region Specific Time Trend	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust
R-squared	0.9913	0.9932	0.9940	0.9940
Number of observations	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or the unemployment pool. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A50: Weighted FE Regression, Year FE, Open Unemployed, Shares

	(1)	(2)	(3)	(4)
ln Digi	0.0788	0.1187*	0.1186*	0.0988
	(0.0679)	(0.0691)	(0.0709)	(0.0681)
ln High Edu		-0.5115**	-0.4288*	-0.3420
		(0.2151)	(0.2569)	(0.2387)
ln Low Edu		0.1483	0.2156	-0.1051
		(0.3700)	(0.3647)	(0.3525)
ln Old Unemployed			0.0475	0.0217
			(0.0420)	(0.0427)
ln Young Unemployed			0.0480	0.0493
			(0.0563)	(0.0545)
ln Female Unemployed			-0.0698	-0.0921
			(0.0968)	(0.0868)
ln Foreign born				-0.2847***
				(0.0719)
Constant	1.5707***	1.1791*	1.5169**	0.5027
	(0.3009)	(0.6213)	(0.6964)	(0.6790)
Year Dummy	Yes	Yes	Yes	Yes
Region Dummy	Yes	Yes	Yes	Yes
Standard Errors	Robust	Robust	Robust	Robust
R-squared	0.9934	0.9936	0.9936	0.9940
Number of observations	1368	1368	1368	1368

Note: The independent variable is matching efficiency, computed as the residual of matching function with open unemployed and seasonal fixed effects. Control variables are expressed as shares of the population or the unemployment pool. All standard errors are robust and clustered at region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Appendix 2

Job Title Codes for IT specialists (Section 5.2)

In the SSYK4 classification IT jobs comprise: 1236 IT Managers, 2131 System architects and Programers, 2139 Other Data Specialists, 3121 Data technicians, 3122 Data operators. The IT jobs in the SSYK4_2012 classification comprise basically the same jobs as in the SSYK4 classification, but provides a higher number of categories allowing for a more granular assessment: 1311 and 1312 IT Managers, 2511 System Analysts and ICT Architects, 2512 Software and System developers, 2513 Games and Digital Media developers, 2514 System testers and test managers, 2512 System administrators, 2516 ICT Security Specialists, 2519 Other ICT Specialists, 3511 ICT operations technicians, 3512 ICT support technicians, 3513 System administrators, 3514 Computer network and system technicians, 3515 webmasters and web administrators. Job titles starting with a 1 are manager positions, starting with a 2 are 'occupations requiring an advanced level of higher education' and starting with a 3 are 'occupations requiring higher education qualifications or equivalent'.

Multi-Year Differences (Section 6.2.3)

Another sensitivity check is to estimate a first difference regression with differences over multiple years. That means computing the differences over time spans of 4 years or more. While this reduces the number of observations, it also entails the opportunity to gain some insights into the correlation coefficients of digitization over the medium and long run. With a declining number of observations region dummies becomes less reasonable. These dummies represent the region specific time trend. Then for difference periods of 9 years the region specific time trend is formally based on two observations (2 periods), for the 19 year (1 period) difference period the region specific time trend would be formally based on one observation. Therefore, I do not include region dummies in the regressions.

Considering longer time periods a mostly positive correlation between digitization and matching efficiency can be observed. Interestingly the magnitude of the coefficient is also very stable and very similar to the previous results in most versions. On a side note, is is interesting how the R^2 increases with longer difference periods, the levels are remarkably high given that no fixed effects are employed. Also the two period regression with two periods 1998-2007 and 2008-2016 has large explanatory power. The financial crisis lies exactly at the beginning of the second period and therefore the regression can treat the second period like a shift in level due to the financial crisis. This confirms my intuition about including Dummies for

Table A51: Long Differences, Open Unemployed, Shares

	Δ 2000 -2004 Δ 2000 -2008 Δ 2008-2012 Δ 2012-2016	Δ 1998 -2002 Δ 2002 -2006 Δ 2006-2010 Δ 2010-2014	Δ 1998 - 2004, Δ 2004 - 2010 Δ 2010-2016	Δ 1998 -2007 Δ 2007 - 2016	Δ 1998 - 2016
D.ln Digi	-0.0261	0.1048**	0.1086*	0.0939	0.0845
	(0.0559)	(0.0467)	(0.0569)	(0.0580)	(0.0688)
D.ln Foreign born	-0.3491***	-0.1936*	-0.2096**	-0.1382	-0.2900**
	(0.0887)	(0.1137)	(0.0988)	(0.1144)	(0.1131)
D.ln Old Unemployed	0.1870***	-0.0317	0.1429***	0.1438**	0.0214
	(0.0495)	(0.0391)	(0.0343)	(0.0565)	(0.0870)
D.ln Young Unemployed	0.2145***	0.1082	0.1524**	0.1317*	0.0141
	(0.0454)	(0.0739)	(0.0638)	(0.0747)	(0.1939)
D.ln Female Unemployed	-0.1526*	0.0640	0.0314	0.1404	0.1652
	(0.0868)	(0.1027)	(0.1122)	(0.1998)	(0.3006)
D.ln High Edu	-1.0778***	0.6362**	-0.7040***	0.1943	-0.2060
	(0.2454)	(0.2591)	(0.1793)	(0.1900)	(0.5299)
D.ln Low Edu	-0.2760	1.6389***	0.4969*	1.5867***	0.8146
	(0.2444)	(0.2973)	(0.2771)	(0.3226)	(0.5718)
R-squared	0.6200	0.4623	0.7271	0.8395	0.9171
Number of observations	288	288	216	144	72

Note: The independent variable is matching efficiency, computed as the residual of matching function of total and open unemployed and seasonal fixed effects. Control variables are population shares. The variables female, young and old unemployed depict the share of the respective unemployment pool. All standard errors are robust and clustered at region level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Year 2008 and Year 2009 in the earlier regressions. However, the results do also seem to depend on the years chosen for the difference calculation. When 4 year differences are computed for the time 2000-2016, the correlation is basically zero, but when 4 year differences are calculated for 1998-2014, the correlation is significantly positive. Other coefficients of control variables also vary greatly between these two specifications. This difference could be due to differential effects in the years 1998 and 2016, or just because there is a lot of volatility within those 4 year periods and choosing different years for computing the difference consequently impacts the results massively. This aspect reduces the relevance of these results.

Variable List and Data Sources (Section 6.2)

The following table lists the variables used in the different regressions in the results part. The variable list is not structured by tables, but by data types. This means that the same variables are used by regressions with different functional forms but with the same data type. The different data types are population shares and absolute values; and total unemployed and openly unemployed. The two tables "FE regression with time trends, Openly Unemployed, Absolute values" and "FD regres-

sion, Openly Unemployed, Absolute values" share the same data type and thereby also the same variable definitions.

Regarding the data sources, SCB stands for Statistics Sweden, AF stands for Arbetsförmedlingen (the Swedish Public Employment Service). All data points which were used in regressions and statistics throughout the paper refer to an aggregate level, either the FA regions or Sweden nation wide. Some parts of the data have been aggregated from anonymized individual level data obtained from SCBs data base LISA. The data processing was conducted on SCBs server MONA and followed the data protection guidelines imposed by Statistics Sweden and the National Institute of Economic Research.

Variable	Definition	Data Source
	Openly Unemployed, Population Shares	
Digi	The number of persons who are working in the FA region and have ICT education, divided by the total number of persons working in the region.	SCB (LISA)
Old Unemp.	Share of the openly unemployed who are aged 55- 64 in the stock of openly unemployed in the region. Yearly Average.	AF
Young Unemp.	Share of the openly unemployed who are aged 18-24 in the stock of openly unemployed in the region. Yearly Average.	AF
Female Unemp.	Share of the openly unemployed who are female in the stock of openly unemployed in the region. Yearly Average.	AF
Foreign Born	The number of persons living in a FA region who were not born in Sweden, divided by the total population number of the FA region.	SCB (LISA)
High Edu	Number of persons living in the FA region with an education level of Sun2000niva 5 and 6, divided by the total population number of the FA region	SCB (LISA)
Low Edu	Number of persons living in the FA region with an education level of Sun2000niva 1 and 2, divided by the total population number of the FA region	SCB (LISA)

Openly Unemployed, Absolute Values		
Digi	Number of persons who are working in the FA region and have ICT education.	SCB (LISA)
Old Unemp.	Number of the openly unemployed who are aged 55- 64 in the region. Yearly Average.	AF
Young Unemp.	Number of the openly unemployed who are aged 18-24 in the region. Yearly Average.	AF
Female Unemp.	Number of openly unemployed who are female in the region. Yearly Average.	AF
Foreign Born	The number of persons living in a FA region who were not born in Sweden.	SCB (LISA)
High Edu	Number of persons living in the FA region with an education level of Sun2000niva 5 and 6.	SCB (LISA)
Low Edu	Number of persons living in the FA region withan education level of Sun2000niva 1 and 2.	SCB (LISA)
Population	Number of people working in the FA region.	LISA
Employment	Number of persons, above age 16, living in the FA region.	LISA

Total Job seekers, Population Shares		
Digi	The number of persons who are working in the FA region and have ICT education, divided by the total number of persons working in the region.	SCB (LISA)
Old Unemp.	Share of the total job seekers who are aged 55-64 in the stock of total job seekers in the region. Yearly Average.	AF
Young Unemp.	Share of the total job seekers who are aged 18-24 in the stock of total job seekers in the region. Yearly Average.	AF
Female Unemp.	Share of the total job seekers who are female in the stock of total job seekers in the region. Yearly Average.	AF
Foreign Born	The number of persons living in a FA region who were not born in Sweden, divided by the total population number of the FA region.	SCB (LISA)

High Edu	Number of persons living in the FA region with an education level of Sun2000niva 5 and 6, divided by the total population number of the FA region	SCB (LISA)
Low Edu	Number of persons living in the FA region with an education level of Sun2000niva 1 and 2, divided by the total population number of the FA region	SCB (LISA)
Unemp in Program	Share of total job seekers who are participating in a labor market program among all total job seekers. That refers to job seekers in the category "Program med aktivitetsstöd" in the Statistics. Yearly Average. Programs encompass Arbetsmarknadsutbildning, Arbetspraktik, Etableringsprogrammet kartläggning, Etableringsprogrammet, Förberedande insatser, Förberedande utbildning, Jobb- och utvecklingsgarantin, Jobbgaranti för ungdomar, Projekt med arbetsmarknadspolitisk inriktning, Stöd till start av näringsverksamhet, Validering.	AF
Employed Job Seekers	Share of total job seekers who are employed and do not receive government benefits among all total job seekers. That refers to job seekers in the category "Arbete utan stöd" in the Statistics. Yearly Average. This encompasses Deltidsarbetslösa, Tillfällig timanställning, Sökande med tillfälligt arbete, Ombytessökande	AF

Total Job seekers, Absolute Values		
Digi	Number of persons who are working in the FA region and have ICT education.	SCB (LISA)
Old Unemp.	Number of the total job seekers who are aged 55- 64 in the region. Yearly Average.	AF
Young Unemp.	Number of the total job seekers who are aged 18-24 in the region. Yearly Average.	AF
Female Unemp.	Number of the total job seekers who are female in the region. Yearly Average.	AF

Foreign Born	The number of persons living in a FA region who were not born in Sweden.	SCB (LISA)
High Edu	Number of persons living in the FA region with an education level of Sun2000niva 5 and 6.	SCB (LISA)
Low Edu	Number of persons living in the FA region withan education level of Sun2000niva 1 and 2.	SCB (LISA)
Unemp in Program	Number of job seekers who are participating in a labor market program in the FA region. That refers to job seekers in the category "Program med aktivitetsstöd" in the Statistics. Yearly Average. Programs encompass Arbetsmarknadsutbildning, Arbetspraktik, Etableringsprogrammet kartläggning, Etableringsprogrammet, Förberedande insatser, Förberedande utbildning, Jobb- och utvecklingsgarantin, Jobbgaranti för ungdomar, Projekt med arbetsmarknadspolitisk inriktning, Stöd till start av näringsverksamhet, Validering.	AF
Employed Job Seekers	Number of job seekers who are employed and do not receive government benefits in the FA region. That refers to job seekers in the category "Arbete utan stöd" in the Statistics. Yearly Average. This encompasses Deltidsarbetslösa, Tillfällig timanställning, Sökande med tillfälligt arbete, Ombytessökande	AF
Population	Number of people working in the FA region.	SCB (LISA)
Employment	Number of persons, above age 16, living in the FA region.	SCB (LISA)