Formal Education, Mismatch and Wages After Transition: Assessing the Impact of Unobserved Heterogeneity using Matching Estimators*

Ana Lamo

European Central Bank (DG-Research)

Julián Messina

World Bank, Universitat de Girona and IZA

May 2010

Abstract

This paper studies the incidence and consequences of the mismatch between formal education and the educational requirements of jobs in Estonia during the years 1997-2003. We find large wage penalties associated with the phenomenon of educational mismatch. Moreover, the incidence and wage penalty of mismatches increase with age. This suggests that structural educational mismatches can occur after fast transition periods. Our results are robust for various methodologies, and more importantly regarding departures from the exogeneity assumptions inherent in the matching estimators used in our analysis.

Keywords: *Education mismatch, Wage determination, Matching Estimators*

JEL Classification: *J0.*

*The opinions expressed in this article do not necessarily reflect the views of the European Central Bank or the World Bank. We are grateful to two anonymous referees, Raul Eamets, Sara de la Rica, Anna Sanz de Galdeano, Marcello Sartarelli, seminar participants at the Bank of Estonia, at the 2008 European Association of Labour Economists and at the 2008 European Economic Association annual meetings for helpful comments. Julián Messina would like to acknowledge the financial support provided by the research grant No. SEJ2007-62500 by the Spanish Ministry of Science and Technology.

1 Introduction

This paper analyzes the incidence and consequences of educational mismatch in Estonia, an economy that has suffered a rapid period of structural transformation. Estonia joined the European Union (EU) together with other 9 Central and Eastern European countries on May 1st 2004. A prominent feature of these new EU members was their low GDP per capita when compared to the older partners. Overall, the new members increased the EU population by 20%, reaching 450 million, while the Union's GDP increased by only 5%. These sharp differences in terms of GDP per capita brought new challenges for both the old and new member states. Regarding the latter group, it is often argued that one of the channels that should facilitate their economic convergence towards the levels of wellbeing experienced by the former states (hereinafter the EU-15) is the high level of education of their workforce.[1] This view is based on indicators of average years of schooling in the new members, which are often higher than those of the EU-15. In Estonia, among people aged over 25 the average level of schooling in 1999 was 9.2 years, clearly greater than the EU-15 average of 8.7 (Barro and Lee, 2001). For the other new EU members, the numbers are similar; e.g., Poland had 9.2 years of schooling on average.

We argue that this fact needs to be qualified, in the sense that in the new EU members, education was designed to meet the needs of a centrally planned economy, and workers' human capital might not be best suited to rapidly catching up with the west. In Lamo, Messina and Wasmer (2006) we show that skill specificities in former centrally planned economies slow down significantly the adjustment process in periods of rapid structural changes. In this paper, we study mismatches in "formal" education (hereafter called educational mismatch) in Estonia during the period 1997-2003. Workers' human capital depreciation might bring in mismatches between their formal education and their jobs. This might not necessarily imply a mismatch between workers' productive skills and the skills required by the market, but simply that workers were trained with skills that are actually not demanded anymore. Hence, even if individuals are optimally allocated given their productive skills, they might be formally overeducated. For the purposes

[1] See for instance Caselli and Tenreyro (2004), who make this point when they discuss the prospects of convergence for Poland.

of our study, Estonia constitutes an ideal laboratory. Estonia quickly adjusted from a centrally planned to a free market economy in the early 1990s. It went through a process of drastic reforms, which resulted in strong sectoral reallocations and a rapid privatization of the public firms. By 1997, the starting period for this study's analysis, Estonia had been transformed into a fully functional market economy. Moreover, its regulatory and labour market institutions offer a very flexible environment for EU standards, characterized by relatively low employment protection and unemployment benefits, and almost non-existent trade unions. Compared to other transition countries, the relatively weak safety nets and extremely flexible wage setting institutions (see Messina and Rõõm, 2009) in Estonia are likely to keep in the market workers who suffered important skill depreciations.

A large volume of empirical literature studies the consequences of mismatches between workers' formal education and their jobs' educational requirements. In all of these studies a wage penalty is associated with the overeducation phenomenon, i.e., workers who are educated for a more qualified job than the one they hold earn less than workers with the same education but holding in a job that requires their qualification level. A major difficulty in interpreting this wage penalty as the causal effect of educational mismatch on wages lies on the treatment of unobserved heterogeneity. Skills unobserved by the econometrician (e.g., low ability) might be correlated with overeducation and wages, biasing the estimated coefficients. Bauer (2002) shows that almost 70% of the wage penalty associated with overeducation dissipates once individual fixed effects are introduced in a panel framework. Although appealing, panel techniques might not solve all selection problems. If more able individuals are more likely to leave the jobs for which they are overeducated, panel estimates would also deliver biased coefficients. Moreover, as is well know, individual effects exacerbate the impact of measurement errors, inherent in any measure of overeducation. Other approaches dealing with unobserved heterogeneity include the use of instrumental variables (Dolton and Silles, 2001), controlling for ability using different proxies (see Chevalier, 2003 and McGuinness, 2003) and proxying ability by differentiating between groups of workers in a quantile regression framework (Budría and Moro-Egido, 2006 and McGuinness and Bennett, 2007). More recently, non-parametric approaches have been proposed to deal with the problem of confounding factors (McGuinness, 2008, McGuinness and Sloane, 2009 and Mavromaras, McGinness and Fok, 2009). Our paper is directly related to the latter.

Our preferred estimates rely on nearest neighbors matching estimator techniques (Abadie and Imbens, 2006), which do not impose any functional form on the impact of "formal" overeducation (hereafter overeducation) on wages. More importantly, we do several robustness exercises that show the impact of unobserved heterogeneity in the estimates, following Ichino, Mealli and Nannicini (2008). In the absence of strong priors about the exogeneity of overeducation or credible instruments, our strategy is to assess non-parametrically how robust our results are to the presence of an unobserved confounder, which can be simulated in several forms. A similar strategy has been proposed by McGuinness and Sloane (2009) and Mavromaras et al. (2009). The former studies the effects of educational mismatches on wages in the market of UK graduates, while the latter studies the incidence and wage impact of skill mismatches in the Australian labor market. In both cases, sensitivity analysis of the wage effects of overeducation (or overskilling) are carried out following Rosenbaum (2002), which allows examining the sensitivity of significance levels and confidence intervals in the presence of an unobserved variable. Like Rosenbaum (2002), we do not rely on any parametric model. Differently from previous papers, our method aims instead at assessing the sensitivity of point estimates (rather than confidence intervals), under different scenarios for the distribution of the confounding factor.

Our findings indicate that the incidence of overeducation in Estonia during the period of study is higher among older workers, and in the case of women it increases monotonically with age. The wage penalty associated with overeducation is quite large, lowering wages on average by 24%. Moreover, this wage penalty also increases with age: older overeducated workers receive a higher wage penalty than (otherwise similarly) younger overeducated workers. In order to assess the impact of various forms of unobserved heterogeneity in the estimated effects, we first assume that selection on the unobservables is the same as selection on the observables, in an empirical strategy that closely resembles Altonji, Elder and Taber (2005). We find that in this case the impact on the estimates is minimal. Next, we pose to the data the following question. Suppose there is an unobserved factor, call it "ability", which is negatively correlated with wages but is more likely to be present among the overeducated individuals. Under what distribution of this confounding factor the impact of overeducation on wages is driven to zero? As expected, our results are sensitive to the distribution of this confounding factor, but remain negative and highly significant even in relatively extreme cases. Suppose that among the overeducated 85% are "low ability" individuals, while only 15% of the well-matched belong to this group. Further, assume that the probability of finding a "low ability" worker with a wage below the mean among well-matched individuals doubles the probability of finding a "low ability" worker with a wage above the mean. Then, once this confounding factor is taken into account in the estimation, the elasticity of wages on overeducation declines to -0.16 (from -0.24 in the baseline) for females, and to -0.14 (from -0.24) for males, and remains highly significant. Similar exercises lead us to conclude that all the wage penalty associated with educational mismatches is unlikely to be driven by unobserved worker characteristics.

The rest of the paper is organized as follows. Section 2 presents an institutional and macroeconomic background of the Estonian economy. Section 3 discusses the theoretical underpinnings of overeducation, the different measures of overeducation proposed in the literature and the data used in this paper, while Section 4 studies the incidence of overeducation in Estonia. Section 5 discusses the matching estimator applied in the rest of the paper and studies the consequences of overeducation for wages. Section 6 assesses the quality and reliability of the estimates, paying special attention to the role of unobserved heterogeneity. Section 7 presents our study's conclusions.

2 Background

Estonia joined the EU following more than a decade of major reforms intended to reallocate its resources and change its institutional structures so that it would be compliant with market economy principles. Estonia is a small country that, after gaining independence from the Soviet Union in 1991, introduced its own currency pegged to the German

DM. It then launched drastic economic reforms, which have been qualified as leading to some of the most rapid and successful transitions. Currently Estonia has an extremely open economy, with a reasonably sized public sector (most public companies were privatized before 1993). This transition to a market economy was accomplished through a large increase in worker flows (Haltiwanger and Vodopivec, 2002) and sectoral reallocation: the proportion of agricultural workers dropped from over 20% in 1990 to 8% in 2000. At the same time, there was a remarkable increase in the proportion workers employed in services: from 43% in 1990 to about 60% in 2002. Following the 1992 reforms, Estonia experienced negative real GDP growth rates during three consecutive years until 1995, when the economy started to recover. In 1999, one year after the announcement that Estonia would join the EU, the GDP growth rate again became negative, and since then recovered strongly in 2000 to reach a stable growth level of 6.5% up to 2003. The unemployment rate was 9.2% in 1998 and increased between 1999 and 2000 (11.3 and 12.5% respectively), but subsequently it declined to 11.8 and 9.1% in 2001 and 2002.

By EU standards, Estonia's market economy is considered very flexible in terms of labour legislation and labour market institutions. Moreover, there is no effective trade union movement influencing wages in Estonia. Since 1991 the government has only set minimum wages, while individual wages have been set at the firm level through bilateral agreements between employers and employees. No policy has been established to prevent bankruptcy, and layoffs and separation costs remained very low during the period of analysis. The "Employment Contracts Act" was introduced in 1992, in order to stimulate labour reallocation. This law gave employers the right to layoff workers with two months notification. At the same time workers are entitled to a maximum severance payment equivalent to 4 times their monthly salary. Because this established no limitations on renewals, it also opened up the possibility of the extensive use of temporary contracts, but their cumulative duration should not exceed 5 years. Unemployment insurance and income support are not very generous in Estonia. Unemployment benefits have been fixed at 60% of the minimum wage, which amounts to less than 25% of the average wage. Replacement ratios dropped from 32% in 1990 to 7% in 1998, and eligibility conditions are also very restrictive. The duration of unemployment benefits is limited to

6 months, after which the unemployed could receive social assistance, which is also very limited. Consequently, it can be hardly argued that unemployment benefits and social assistance have any disincentive effect on labour supply. Only training programs act as active labour market policies in Estonia. From 1993 to 1995, both public expenditures and participation in training programs increased substantially.

3 Educational mismatch: theoretical considerations and empirical identification

Educational mismatch may be a permanent or a temporary state. In a matching framework, Jovanovic (1979) shows that temporary educational mismatches can arise due to inefficiencies in the functioning of the labour market associated with imperfect information and mobility. Educational mismatch is also a temporary phenomenon in a career mobility framework, where young workers voluntarily accept jobs for which they are overeducated in exchange of a skill acquisition process that will complement their qualifications (Sicherman and Galor, 1990). In both cases workers are expected to improve over time their matches either through job-to-job mobility or mobility within the firm. Hence, educational mismatches are expected to alleviate while the worker obtains labour market experience. Educational mismatches, on the other hand, might reflect a permanent phenomenon. This is the case in models where employers use formal education as a screening device (Spence, 1973) or when low and high educated workers compete for scarce jobs in the presence of frictions (see Albrecht and Vroman, 2002 and Dolado, Jansen and Jimeno, 2008). In the latter models, structural mismatches may be exacerbated by supply forces such as rapid educational upgrades in the labour force, or by demand forces, such as skill biased technological change. Both cases imply a rapid change in the demand for or supply of certain educational group that cannot be easily matched on the other side of the market.

A large volume of empirical literature studies the determinants and economic consequences of mismatches between individuals' formal education and the educational requirements of their jobs. Following Freeman's (1976) seminal book, this literature identifies workers as being over or undereducated relative to their job, and studies the consequences of educational mismatches on wages and other labour market outcomes. Regarding the impact of overeducation on earnings, McGuinness (2006) reviews this literature and concludes that all studies find a significant wage penalty associated with overeducation, which averages at 15 percent lower wages. Although there is some disagreement in the literature on the rationale for temporary mismatches, i.e., if they are related to market failures or career mobility, most studies finds signs of overeducation being a temporary phenomenon. The fact that overeducation is typically higher during the school to work transition years is suggestive of the short term nature of educational mismatches. Consistent with this interpretation, overeducation has been found to be associated with higher mobility rates (e.g., Alba-Ramírez, 1993, Sloane, Battu, and Seaman, 1999 and Groot and van den Brink, 2003) or within firm promotions (e.g., Alba-Ramírez and Blázquez, 2003 and Groeneveld and Hartog, 2004). Also in line with a temporary interpretation, Groot (1996) finds that overeducation declines with age. However, in contrast with overeducation being a temporary phenomenon, several studies of the labour market of graduates fail to find a declining incidence of overeducation during the first years after graduation (see McGuinness and Wooden, 2009 for a discussion).

The empirical literature on education mismatch typically relies on three kinds of measures of over/undereducation, depending on the specific features of the data set and the information available: i) objective or data-based measures are based on contrasts between the actual distribution of workers' educational attainment and an (estimated) adequate level of education per occupation. This adequate level of education is either measured as a function of the average (see Verdugo and Verdugo, 1989) or the modal (Mendes de Oliveira, Santos and Kiker, 2000) level of education for each occupation, ii) measures based on the contrast between the educational level of workers and the required level of education for their job, derived from systematic evaluation by job analysts who specify the required level of education for the job titles in an occupational classification (Hartog, 1980) and iii) the so-called subjective or direct measures of education mismatch, based on workers' self-assessments (Sicherman, 1981). There are pros and cons associated with each of these measures. The advantage of educational mismatch measures based on worker self-assessment is that they identify how the individual's situation can be assigned relative to the education mismatch, and precisely with the individual's job, and not with any kind of aggregate (Hartog, 2000). The disadvantage however is that they might suffer from workers' misperceptions regarding their actual job requirements. The main arguments in favour of data-based indexes is that they are not impeded by subjectivity, yet one important drawback is that different definitions of what constitutes adequate schooling levels typically deliver very different results. Conceptually, systematic analysis from job analysts should provide the best measures of required education for the job. However, van der Velden and van Smoorenburg (1997) show that worker self assessments of the educational requirements for their jobs are typically much more accurate than expert rating of job titles. Moreover, systematic expert rating analysis is expensive and consequently infrequently updated. This renders the characterization of jobs rapidly obsolete, especially in periods of rapid structural changes.

3.1 The data

The Estonian labour force survey (ELFS) has a structure very similar to the LFSs carried out in the other EU member states, using internationally agreed concepts and definitions as proposed by the International Labor Organization (ILO). It contains standard demographic and job characteristics, and its longitudinal nature allows individuals to be followed for a maximum of 1.5 years. The ELFS was first conducted in 1995 and on an annual basis until 2000q1, when the methodology changed. From 2000q2 the data was collected quarterly and the panel followed a 2-2-2 rotation plan. This implies that every household was interviewed for two quarters, not observed during the next two quarters and interviewed again for two consecutive quarters. All the surveys are stratified samples of the population census, and are representative of the population in the age bracket 15-74. The response rate is very high, always above 90%. The 1997 survey interviewed 5,555 individuals, while the 1998 and 1999 ELFS sampled around 14,000 individuals, and 25% of the 1998 sample was retained in the 1999 survey. After 2000, some 10,000 individuals are interviewed every year. From the first part of the survey we retained the 1997, 1998 and 1999 waves, which contain information from the second quarter of the year, and then we exploited the quarterly information thereafter.[2] Our analysis period is 1997-2003.

The Estonian labour force survey made it possible to construct a direct educational mismatch measure, based on the respondents' perceived level of required education for their job. All employees were asked: "Does your job correspond to your educational level?" and were offered three response options: "Yes", "No, the job presupposes a more advanced level of education" or "No, the job presupposes a lower level of education". Note that the question explicitly asks about the educational level, and not about the skills of the individual. The question should be informative about the quality of the match between the formal education received by the worker and the level of education required for her job. Hence, it is certainly possible that workers who answer negatively to the above question are paid according to their relative productivities. In other words, the presence of mismatch in our sample is not necessarily indicative of market failures in the Estonian labour market. It might simply reflect that workers were trained for jobs or occupations that are not demanded in the market anymore.

Using this information it is possible to directly construct measures of over/undereducation. In our data 12.6% of the workers declared that they were overeducated for the job, and 2.5% were undereducated. Since the estimation methodology described in Section 5 was designed for dichotomous treatment, we focused the analysis on the overeducated, and excluded undereducated from the sample.[3] The outcome measure in the paper is the net monthly wage. When the individual is employed in more than one job, we retain the job with the highest salary as the main job. We limit the sample to employees, in the age bracket 16-65.

[2] The 1997, 1998, and 1999 contain retrospective information that can be used to construct quarterly data. However, the information on the matching between the education of the individual and the job refers to the reference week (the second quarter of each year).

[3] We have estimated OLS regressions, including both over and undereducation indicators, and did not find a significant difference in wages between undereducated and well-matched workers. We obtain very similar results when we apply matching estimators to the sample of workers who are undereducated, having well-matched workers as the control group.

4 Who is overeducated in Estonia

Table 1 lists certain summary statistics showing important differences between overeducated and well-matched workers. On average the latter earn more and have fewer years of education than the overeducated. Among women and interestingly among older workers, there is a higher incidence of the overeducation. For more detail on the relationship between overeducation and age see Figure 1, which reveals that the incidence of overeducation for females increases monotonically with age, yet slightly less so for older workers. For males, the age overeducation profile increases (non-monotonically) with age, and among the older cohorts overeducated individuals are overrepresented. It is also interesting to note that Estonian origin individuals (based on the first language spoken at home) are much more likely to be well match than workers from other origins. As for job features, overeducated workers seem to concentrate more on the private sector and in manufacturing in particular, and also exhibit lower job tenure (4.5 years on average versus 7.3 for the well-matched).

To obtain a better understanding of the factors behind overeducation, we estimate a probit model where the dependent variable is assigned the value 1 when individuals declare themselves to be overeducated for the job. We do this separately for males and females, as overeducation seemed to follow different patterns in each case. Table 2 lists the marginal effects evaluated at the mean of the continuous variables, and discrete changes in the case of the dichotomous variables, of the expected changes in the predicted overeducation probability as a function of personal and job characteristics. Since we observe individuals more than once, we report robust standard errors and allow for clustering at the individual level.

According to the estimates in Table 2, the patterns observed concerning age are confirmed by the regression analysis; i.e., overeducation increases monotonically with age among women, while for males, overeducation is concentrated among the oldest, even though the age profile is not monotonic. Remarkable differences are also observed between the genders concerning the importance of certain job features, such as sector of operation and firm size. Similarly, working in the public sector increases the probability of being overeducated much more for male than for female workers. Differences across

Estonian and other origin remain, but the marginal effects both for males and females become quite small when controlling for other characteristics.

It must be stressed that some of the job characteristics included in these probit regressions might be considered endogenous with respect to overeducation, since overeducated workers may tend to concentrate on certain sectors or remain in their jobs for shorter periods. We will take care of this simultaneity in the next section, when we evaluate the consequences of overeducation for wages.

This exploratory analysis provided a remarkable message in that age increases the likelihood of being overeducated, and this finding is in contrast with previous empirical evidence (see Groot, 1996). From a human capital perspective, this is suggestive of structural changes in the Estonian labour market wherein new abilities are required, that the old educational system failed to provide.[4] The rest of this paper studies the impact of overeducation on wages, and whether that impact differs across cohorts.

5 The impact of overeducation on wages

5.1 Methodology

In this section we investigate the consequences of educational mismatch on wages by estimating Mincerian earnings regressions that include a dummy for overeducation among the covariates, making it possible to compare wages of workers suffering from education mismatch with those of workers having similar features but being well-matched. This approach was first applied by Duncan and Hoffman (1981), and has generated a wide range of literature, typically finding that a wage penalty is associated with overeducation. Our dependent variable is the log of hourly wages and we separate the male and female sub-samples. Moreover, as we observed that the incidence of overeducation was higher among older workers, we split the sample in different age groups to assess whether

[4] In alternative specifications, we have identified individuals who attended school entirely under central planning and compared them to those who got at least some formal education after 1991. The results, available upon request, are very similar for females, suggesting that the likelihood of being overeducated is higher for those who studied during the previous regime. In the case of males, the signs are the expected but significance greatly depends on the specific assumptions we make about the threshold for education before and after.

in the case of Estonia the wage penalty typically associated with overeducation varies with age.

We estimate the wage penalty associated with being overeducated (educational mismatch) using standard regression analysis as well as matching estimators as proposed by Abadie and Imbens (2006), hereinafter referred to as AI. In contrast with OLS, matching does not impose any functional form. The principle behind the simple matching estimator is that for each individual there are two potential outcomes, one for individuals who follow the treatment and another for those who do not follow, i.e., those belonging to the control group. The difference between these potential outcomes lies in the treatment effect on the individual. Only one of these potential outcomes is observed however, and the other needs to be estimated. To do so, the simple matching estimate uses information on similar individuals who follow the opposite treatment. In our case the outcome are wages, the treatment group consists of overeducated workers and the control group comprises well-matched workers. Our objective is to estimate the average effect of the treatment, i.e., the wage penalty for being overeducated.

For each individual i, $i = 1, ...N$, we observe the triple (W_i, X_i, Y_i), where X_i is a vector of covariates, $W_i \in \{0, 1\}$ is an indicator on whether individual i received treatment or not, and Y_i denotes the realized outcome, which is equal to $Y_i(0)$ if the individual is well-matched, i.e., she is part of the control group and to $Y_i(1)$ if she is overeducated, i.e., she belongs to the treatment group.

$$Y_i \equiv Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1. \end{cases} \tag{1}$$

We are interested in what AI refers to as the "average effect on the treated population" (ATT) ($\tau^{p,t}$) and in the "sample average treatment effect on the treated" ($\tau^{s,t}$):

$$\tau^{p,t} = E[Y_i(1) - Y_i(0)|W_i = 1] \text{ and } \tau^{s,t} = \frac{1}{N_1} \sum_{i:W_i=1} (Y_i(1) - Y_i(0)),$$

where $N_1 = \sum_{i=1}^{N} W_i$ stands for the number individuals in the treated group.

To ensure identification and consistency for the estimated treatment effects, two regularity conditions must hold:

Unconfoundedness: for almost every x in the support of X, the assignment to treatment W is independent of the outcome, conditional on the covariates X;

$$E[Y(w)|X = x] = E[Y(w)|W = w, X = x]. \tag{2}$$

This is also known as selection of observables or the conditional independence assumption. This assumption is crucial, as it allows the realized outcome of individuals having the same covariates values as the opposite group to be used as a valid control group. Thus, the average treatment effect can be recovered by averaging $E[Y|W = 1, X = x] - E[Y|W = 0, X = x]$ over the distribution of X.

Overlap: for almost every x in the support of X, $c < Pr(W = 1|X = x) < 1 - c$, for some c. This assumption implies that the conditional probability of receiving treatment, also known as the propensity score, is bounded away from zero and one. This simply guarantees that for any treated individual there would be some individuals in the non-treated group having similar covariate patterns.

For a formal discussion of these regularity conditions see Abadie and Imbens (2006) and Rosenbaum and Rubin (1983). Unconfoundedness is the most controversial assumption in most empirical applications, and our case constitutes no exception. We discuss this thoroughly in section 6 and provide evidence supporting its plausibility.

The matching estimator that we consider imputes the missing potential outcome for an individual, using information on observed outcomes of individuals who are "close" in terms of their covariate values. We will do matching with replacement, i.e., allow each individual in the control group to be used in more than one match, since this technique produces better matches than that without replacement by increasing the set of possible matches. The simple matching estimator for the ATT estimates the missing potential outcomes $Y(0)$ when $W_i = 1$ as the average of the outcomes of the nearest neighbors belonging to the control group:

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \frac{1}{M} \sum_{j \in I_M(i)} Y_j & \text{if } W_i = 1 \end{cases}, \tag{3}$$

where $I_M(i)$ is the set of indices for the first M matches for individual i. Hence, the simple matching estimator for the average treatment effect for the treated discussed in AI is:

$$\hat{\tau}^{sm,t} = \frac{1}{N_1} \sum_{i:W_i=1} (Y_i - \hat{Y}_i(0)), \tag{4}$$

where N_1 denotes the number of treated individuals in the sample. AI show that due to matching discrepancies this estimator has a bias of the order $O(N^{-1/K})$, where K is the number of continuous covariates. They suggest combining the matching process with a regression in order to adjust the differences within the matches to the differences in their covariate values. This adjustment is based on an estimate of the regression function $\mu_w(x) \equiv E[Y(w)|X = x]$ for the control group.[5] This bias adjustment makes the matching estimators $N^{1/2}$ consistent. In our case, as will be seen in the next section, no major discrepancies exist between simple matching and bias corrected estimators. This might be expected because our only continuous control variables are age, job tenure and years of schooling.

5.2 Estimation Results

Our aim is to provide robust evidence of the consequences of educational mismatch. To this end we report average wage penalties for the overeducated according to various estimation methods: i) the unconditional mean difference estimator, ii) OLS estimators and iii) several matching estimators. We present the results of simple matching and biased adjusted matching techniques for one and four matches (M=1 and M=4, respectively). We also examine two different equation specifications; the first including a restricted number of controls, from which we have excluded certain job features that could be endogenous to overeducation (e.g., tenure, firm size, sector of operation). These are most likely intermediate outcomes, and hence if included in the regression a downward bias is likely to influence the overall effect of overeducation on wages. The second specification includes a larger number of controls, some of which are the above mentioned potential

[5] AI use nonparametric estimation to impute the value for the non-treated.

intermediate outcomes.

Table 3 shows the estimated average wage penalty (ATT) separately for the entire sample of males and females, using the above- mentioned estimation methods and the baseline set of covariates. The regressors included are a dummy for ethnic origin, two dummies for marital status, years of education, a quadratic in age, time and regional dummies and an indicator variable for overeducation. The results are very robust across the estimation methods. For Estonia the average wage penalty due to overeducation is about 24 to 27 % for females, depending on the estimation method, and slightly lower at 18 to 24% for males.[6] It should be noted that this wage penalty is quite high when compared to available results for other European countries (see Groot, Maasen and Brink, 2000). Using comparable data and an overeducation measure similar to ours, Budría and Moro-Egido (2006) find that the wage penalty associated with overeducation ranges from 2.6 to 10.9% across 12 EU-15 countries in the period 1994-2001.

Table 4 lists results for various age groups; it decomposes the sample into four age categories: 16-29, 30-39, 40-49 and 50-64. When we look at age groups, it is interesting to note that the penalty for younger cohorts (aged 16 to 29) is drastically smaller and less stable across the various methods; between 4 and 9% for women and between 8 and 13% for men. In both the male and female cases the wage penalty associated with overeducation increases with age. In the case of females this increase is progressive: overeducated females aged 50-64 have a higher wage penalty than middle aged females (40-49), with a difference of around 5 percentage points and with small variations, depending on the estimation method used. The difference between middle-aged females and those aged 30-39 is slightly smaller, at about 3 percentage points. For males the differences across cohorts have similar magnitudes; with the oldest males hit by the highest wage penalty, ranging between 33 and 35%.[7]

[6] A potential problem we do not deal with here relates to the labour market participation decision of females. While female labour force participation is relatively high in Estonia (averaging at 66% during the sample period) it is clearly lower than males (at 78%). To the extent that the most severely overeducated females were out of the labour market, our estimates of the effects of overeducation on female wages would be downward biased.

[7] We have also splitted the sample between workers who finished school before 1991 and those who at least had some schooling after this year, and we obtain very similar results. While the estimated effects of overeducation on wages are negative and highly significant in the case of workers trained entirely during

Table 5 lists the estimated ATT according to age groups for our extended specification, where we add to the above-mentioned control set the following (potentially endogenous) variables: a quadratic in tenure, a public sector dummy, firm-size dummies and sectoral dummies. The table only displays OLS and our preferred matching estimator (the bias adjusted matching estimator for one match), because the results are very similar for the alternative matching methods. Drastic changes are not observed with respect to the basic specification. The order of magnitude of the wage penalty as well as the age profile and the comparison between males and females barely remain unchanged, but there is a slightly lower wage penalty for the oldest group in both genders, as well as for males aged 30-39.

On average, wage penalties due to overeducation appear to be much higher in Estonia than in other EU-15 countries where similar studies were done. Interestingly, the wage penalty in Estonia associated with overeducation among young workers is lower than that found for older cohorts, and of similar magnitude to that found in other EU-15 countries. This highlights that the differential behavior of the Estonian labour market when it comes to wage penalties associated with overeducation lies in the older cohorts.

6 Assessing the quality and reliability of the estimates

This section provides some evidence supporting the reliability of our estimates. First, it assesses the quality of matching, that is, whether individuals in the treatment and control groups are really alike. Second, some sensitivity analyses are made regarding the robustness of our estimates in the event that the unconfoundedness assumption fails.

6.1 Quality of the matching

To evaluate the quality of matched pairs used in our estimation we follow the same strategy as Abadie and Imbens (2006). Table 6 lists evidence of the quality of matching for the variables used in the basic specification (excluding potential intermediate outcomes).

the communist regime, they present much smaller orders of magnitude (and are not always statistically different from zero) for workers who obtained their diplomas after 1991.

All covariates were normalized such that their mean would be zero and their variance would be equal to one. The first panel lists the results of the female samples, and the second panel lists those of the male samples. The second and third columns in Table 6 show the average value of each covariate for the overeducated and the well-matched before matching. The difference between the second and the third columns is reported in the fourth column. The fifth and sixth columns list the average of the covariates for both groups, computed using the same observations as those in the single matching case ($M = 1$). The seventh column displays the average difference within the matched pairs for each covariate. The matching is quite good, and its impact on the difference between overeducated and well-matched samples is substantial. Before matching, there were large differences between treated and control units for a relatively large set of control variables (e.g., ethnic origin, divorced/widowed, county dummies). In all cases the average difference between the treatment and control group was much smaller after matching than before (compare columns 4 and 7). For several covariates the matching was even exact (the difference after matching is zero or very close to zero).

6.2 Sensitivity to departures from the unconfoundedness assumption

The main behavioral assumption behind unconfoundedness is that in the case of no treatment the potential outcome $Y(0)$ does not influence the treatment assignment once we condition on the workers' observable features. This assumption is formally untestable, because the available data provides no information regarding the wage distribution for the overeducated workers in the case they were well matched ($Y_i(0)$ when $W_i = 1$), but by using certain additional evidence its credibility can be supported/rejected. The data used in our analysis includes information on a large number of worker and job characteristics collected using the same sample and questionnaire for overeducated and well-matched (treated and not treated) workers. Nevertheless, unobserved workers' heterogeneity and/or measurement errors might be important factors that influence the treatment assignment. Low ability workers might need extra years of education to perform well their jobs. Similarly, discouraged workers might be more inclined to answer that they feel properly suited to a more demanding job. If the market were to attribute a wage penalty for low ability or discouraged workers we would be overestimating the impact of educational mismatch on wages.

To assess whether and to what extent the estimated wage penalty associated with overeducation is robust for a potential unobservable confounder we follow Ichino, Mealli and Nannicini (2008). They propose a sensitivity analysis that builds on Rosenbaum and Rubin (1983) and is based on the following idea: suppose that unconfoundedness is not satisfied given the observables,

$$E[Y(0)|W = 0, X] \neq E\left[Y(0)|W = 1, X\right], \tag{5}$$

but that it would be satisfied if we could observe an additional (unobservable) variable, denoted by U, such that

$$E[Y(0)|W = 0, X, U] = E[Y(0)|W = 1, X, U]. \tag{6}$$

This potential confounder can be then simulated in the data and used as an additional covariate in the estimation. The distribution of the simulated variable can be constructed to capture different hypotheses regarding the failure of the unconfoundedness assumption. The comparison of the estimates obtained with and without the simulated confounder shows to what extent the results are robust regarding the assumption's failure.

We assume that U is a binary variable conditionally independent with respect to the observables and we characterize its distribution by selecting the following probabilities:

$$p_{ij} \equiv \Pr(U = 1|W = i, Y = j) = \Pr(U = 1|W = i, Y = j, X), \tag{7}$$

where p_{ij} is the probability of $U = 1$ if the treatment is i and the outcome equals j, with $i, j \in \{0, 1\}$. Hence, there are four probabilities p_{ij}, one for each one of the groups defined by treatment and outcome. Note that by stating that $i, j \in \{0, 1\}$ we assume that both the treatment and outcome are binary variables, but in our case the outcome (wage) is continuous. We therefore need to discretize it, and we do so by defining a binary variable that takes value one if the wage is lower than the average wage and takes value zero if it is higher. We call this binary variable *outcome.*[8] Then, given p_{ij}, a value of U is attributed to each of the individuals, depending on which of the four groups defined by the treatment status and the *outcome* value she/he belongs to. The simulated U is then treated as any other observed covariate. Given the probabilities p_{ij} we repeat the matching estimation 100 times for each set of values of the variable U and then calculate a simulated estimate of the ATT by averaging over of the 100 estimated ATTs. As such, the sensitivity analysis provides a point estimate of the ATT that is robust for the failure of the unconfoundedness assumption according to that particular configuration of p_{ij}.

We can change the assumptions about the distribution of U, and in this way we can assess the robustness of the ATT with respect to different hypotheses regarding the nature of the confounding factor. An approach related to ours is proposed by Rosenbaum (2002) and has been implemented by DiPrete and Gangl (2004). This method has been applied in the context of overeducation and overskilling by McGuinness and Sloane (2009) and Mavromaras et al (2009). The methodology relies on only one sensitivity parameter (which represents the association between W and U), instead of the four (sets of) parameters specified here. As a consequence, the joint distributions of $W, Y_1, U \mid X$ and $W, Y_0, U \mid X$ are partially identified, and only bounds for significance levels and confidence intervals can be derived. The approach taken here, by including (different sets) of U in the matching set allows us assessing under which particular conditions the estimated ATT (rather than the confidence interval of the original estimates) is driven to zero.

We start presenting a neutral configuration where p_{ij} takes the value 0.5 for every i and j. In this case the distribution of U has no effect on the selection of treatment ($p_{1\cdot} - p_{0\cdot} = 0$) or on the outcome of the non-treated ($p_{01} - p_{00} = 0$).[9] The second configuration considered is such that the distribution of U resembles the age distribution in our sample. Remember that combining the outcome (a wage above or below the mean)

[8] Ichino, Mealli and Nannicini (2006) present two Monte Carlo exercises showing that discretisation assumptions of this kind do not critically affect the results of the sensitivity analysis.

[9] The difference $p_{1\cdot} - p_{0\cdot}$captures a selection effect, since it measures the effect of U on the selection into the treatment. The difference $p_{01} - p_{00}$ might be labeled as an outcome effect, as it captures the effects of U in absence of treatment. Note however, that these effects need to be avaluated after conditioning on W, as shown by Ichino, Mealli and Nannicini (2008).

and the treatment status we have four categories of workers (per gender). Within each of these categories, we calculate the percentage of workers older than the average in the sample, and assign a probability of $U = 1$ equal to this percentage. For example, in the group "overeducated with earnings less than average" pertaining to the female sample, 54% of the workers are older than average. Hence, we assign a value $U = 1$ to 54% of the workers in this group ($p_{11} = .54$ for females). The results of these two configurations are listed in the second and third rows of Tables 7 and 8, for females and males respectively. The first row shows the estimated baseline ATT using only observables ($M = 1$, bias-corrected estimators, see also Table 3) as covariates in the baseline specification. When in the estimation we include the new simulated covariate described above to be neutral (second row), the estimated ATT hardly differs from that of the baseline, the difference being only -0.4 and 0.3 percentage points, for females and males respectively. When the new covariate (simulated confounder) follows the empirical age distribution (third row), the estimated ATTs are again very close to the baseline.

For simplicity in the rest of the exposition, we will think of U as an indicator of low ability, but it could be any confounder factor that is negatively correlated with wages and positively correlated with overeducation. Hence, we will refer to individuals with $U = 1$ as "low ability" individuals, while individuals with $U = 0$ as "high ability". Our next exercises explore how extreme the distribution of U needs to be in order to generate estimates that substantially depart from the baseline. We make two assumptions. First, we postulate that "low ability" individuals ($U = 1$) are over-represented among the treated (overeducated), which amounts to assuming that $p_{1\cdot} > p_{0\cdot}$. Second, we assume that within each group (treated and non-treated) the workers having "low ability" are more likely to obtain lower wages than those with $U = 0$. This implies that $p_{11} > p_{10}$ and $p_{01} > p_{00}$. The results are listed under cases a) to g) in Tables 7 and 8. Concentrating first on the results for females (Table 7), case a) displays strong selection effects and only weak outcome effects ($p_{11} = 0.9$; $p_{10} = 0.8$; $p_{01} = 0.2$; $p_{00} = 0.1$). The estimated impact is reduced from -0.23 (no confounder) to -0.16, but remains highly significant at the 1% level. The following rows reflect a progressive reduction in the share of "low ability" individuals among the overeducated who have higher than average wages, and results are not greatly affected. In case d), which looks at $(p_{11} = 0.9;\ p_{10} = 0.5;\ p_{01} = 0.2;\ p_{00} = 0.1)$, the estimated coefficient is even larger, at -0.17, and highly significant. The most stringent tests are in cases e) to g), where "low ability" individuals are strongly represented among the treated and those obtaining lower than average wages. Even in case g), where we assume that 90% of the overeducated and 40% of the well-matched who suffer a wage penalty are "low ability" workers $(p_{11} = 0.9;\ p_{10} = 0.5;\ p_{01} = 0.4;\ p_{00} = 0.1)$, there is a sizable and statistically significant negative impact of educational mismatch on wages: -0.10 (s.e. 0.019). On top of the strong selection and outcome effects we need to make the extreme assumption that nobody among the well-matched workers with a wage above the mean belongs to the "low ability" group (p_{00}=0) in order to find an estimated impact close to zero. This is so for case g) $(p_{11} = 0.9;\ p_{10} = 0.5;\ p_{01} = 0.4;\ p_{00} = 0)$, where the average coefficient is -0.03 (s.e. 0.017). Similar conclusions are reached with the male sample, presented in Table 8. In sum, to move the ATT away from the baseline results we need to make fairly extreme assumptions regarding the selection effects of U. This bring us to conclude that it is unlikely that unobserved heterogeneity is driving the main results presented in the paper.[10]

7 Conclusions

Estonia has undergone a rapid transition from a centrally planned to a market economy, and then later a rapid transformation in its productive structure as a consequence of its EU accession. This is an ongoing process and the consequences are likely to be long lasting. This paper documents one of the outcomes of such a process: the mismatch between the formal education of workers and the curricular content of their jobs.

Our research finds that there is a relatively high prevalence of educational mismatch in Estonia; more than 12% of workers are formally overeducated for their jobs. More

[10] Ichino, Mealli and Nannicini (2008) show that non parametric bounds for the ATT as those proposed by Manski (1990) have an equivalent in terms of the distribution of U. The assumptions concerning the confounder U that will lead the ATT to the bounds are quite extreme and highly implausible, explaining why non parametric bounds are often uninformative. For the lower bound, we need to assume that among the treated there are only individuals with $U = 1$, i.e. $p_{11} = p_{10} = 1$, and among the well matched all the less able suffer a wage penalty, i.e. $p_{01} = 1$. The upper bound is instead constructed as $p_{11} = p_{10} = 1$ and $p_{01} = 0$. The bounds of the ATT are (-0.56, 0.14) for females and (-0.56, 0.19) for males.

importantly, the incidence and wage penalty for being overeducated in Estonia increases with age, and is most concentrated among those workers who attended school during the centrally planned regime. The wage penalties associated with overeducation are fairly significant (around 26%), except among younger cohorts, a group in which wage losses associated with overeducation are of a magnitude comparable to those found in other European countries. A battery of robustness checks using non-parametric methods suggests that it is unlikely that these results are driven by unobserved heterogeneity.

Our findings are consistent with the expectations from a rapidly changing transition economy. A fast speed of structural change can render obsolete educational diplomas that were obtained in the previous regime, triggering a mismatch between formal education and labour demand. This implies that summary indicators of average years of schooling in transition countries should be treated with caution, since they might constitute a poor proxy for the true human capital of the working age population.

References

[1] Abadie, A. and G. Imbens (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects", Econometrica, 74 (1), 235-267.

[2] Alba-Ramírez, A. (1993), "Mismatch in the Spanish Labor Market: Overeducation?", Journal of Human Resources 28(2), 259-278.

[3] Alba-Ramírez, A. and M. Blázquez (2003), "Types of Job Match, Overeducation and Labour Mobility in Spain", in Overeducation in Europe: Current Issues in Theory and Policy, F. Buchel, A. de Grip, A. Mertens, eds., Cheltenham, UK.

[4] Albrecht, J. and S. Vroman (2002), "A Matching Model with Endogenous Skill Requirements", International Economic Review 43(1), 283-305.

[5] Altonji, J., T. E. Elder and C. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", Journal of Political Economy 113(1), 151-184.

[6] Barro, R. and J. W. Lee (2001), "International Data On Educational Attainment: Updates And Implications", Oxford Economic Papers, 53, 541-563.

[7] Bauer, T. (2002), "Educational Mismatch and Wages: a Panel Analysis", Economics of Education Review 21(3), 221–229.

[8] Budría, S. and A. I. Moro-Egido (2006), "Overeducation and Wages in Europe: Evidence from Quantile Regression", Working Papers EEE-FEDEA 229.

[9] Caselli, F. and S. Tenreyro (2004), "Is Poland the Next Spain?", NBER International Seminar on Macroeconomics, in Clarida, Frankel and Giavazzi (editors), MIT Press.

[10] Chevalier, A. (2003), "Measuring Mismatch", Economica 70(279), 509-531.

[11] DiPrete, T.A. and M. Gangl (2004), "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Mathing Estimators and Instrumental Variable Estimation with Imperfect Instruments", Sociological Methodology, 34(1), 271-310.

[12] Dolado, J. J., M. Jansen and J.F. Jimeno (2008), "On the job Search in a Matching Model with Heterogeneous Jobs and Workers", Economic Journal 119(534), 200-228.

[13] Dolton, P. and M. Silles (2001), "Mismatch in the Graduate Labour Market: Some Evidence from Alumni Data", Centre for the Economics of Education, Discussion Paper Series, 9.

[14] Duncan, G., and S. D. Hoffman (1981), "The Incidence and Wage Effects of Overeducation". Economics of Education Review, 1(1), 75–86.

[15] Freeman, R.B. (1976), "The Overeducated American", Academic Press: New York.

[16] Groot, W. (1996), "The Incidence of, and Returns to Overeducation in the UK", Applied Economics, 28(10), 1345–1350.

[17] Groot, W., and H. Maassen van den Brink (2000), "Over-education in the Labor Market: a Meta-analysis", Economics of Education Review, 19(2), 149–158.

[18] Haltiwanger, J. and M. Vodopivec (2002) "Gross Worker and Jobs Flows in a Transition Economy: An Analysis of Estonia.", Labour Economics, 9(5), 601-630.

[19] Hartog, J. (2000), "Over-education and Earnings. Where are we, Where Should we Go?", Economics of Education Review, 19(2), 131–48.

[20] —— (1980), "Earnings and Capability Requirements", Review of Economics and Statistics, 62(2), 230–240.

[21] Ichino, A, F. Mealli and T. Nannicini (2008), "From Temporary Help Jobs to Permanent Employment: What Can We Learn From Matching Estimators and Their Sensitivity?", Journal of Applied Econometrics, 23(3), 305-327.

[22] Jovanovic, B. (1979), "Job Matching and the Theory of Turnover", Journal of Political Economy, 87(5), 972–90.

[23] Lamo, A., J. Messina and E. Wasmer (2006), "Are Specific Skills and Obstacle to Labour Market Adjustment?", CEPR Discussion Papers 5503.

[24] Mavromaras, K., S. McGuinness and Y. Fok (2009), "Assessing the Incidence and Wage Effects of Overskilling in the Australian Labour Market", Economic Record 85(268), 60-72.

[25] McGuinness, S. (2003), "Graduate Overeducation as a Sheepskin Effect: Evidence from Northern Ireland", Applied Economics 35(5), 597-608.

[26] —— (2006), "Overeducation in the Labour Market", Journal of Economic Surveys, 20(3), 387–418.

[27] —— (2008), "How Biased are the Wage Impacts of Overeducation? A Propensity Score Matching Approach", Applied Economics Letters, 15(2), 145–149.

[28] McGuinness, S. and J. Bennett (2007), "Overeducation and the Graduate Labour Market: a Quantile Regression Approach", Economics of Education Review, 26(5), 521-531.

[29] McGuinness, S. and M. Wooden (2009),"Overskilling, Job Insecurity and Career Mobility", Industrial Relations 48(2), 265-285.

[30] McGuinness, S and P. Sloane (2009), "Labour Market Mismatch among UK Graduates: an Analysis Using Reflex Data", IZA Discussion paper 4168.

[31] Mendes de Oliveira, M., M. C. Santos, and B. F. Kiker (2000), "The Role of Human Capital and Technological Change in Over-education", Economics of Education Review, 19(2), 199–206.

[32] Messina, J, and T. Rõõm (2009), "Downward Wage Rigidity During the Current Economic Crisis", Wage Dynamics Network manuscript.

[33] Rosenbaum, P.R. (2002), "Observational Studies" 2d ed. New York. Springer.

[34] Rosenbaum, P. and D. Rubin (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome", Journal of the Royal Statistical Society. Series B, 45(2): 212–218.

[35] Sicherman, N. and O. Galor (1990), "A Theory of Career Mobility", Journal of Political Economy, 98(1), 169–92.

[36] Sloane, P., H. Battu, and P. T. Seaman (1999), "Overeducation, Undereducation and the British Labour Market", Applied Economics 31, 1437–1453.

[37] Spence, M. (1973), "Job Market Signaling", The Quarterly Journal of Economics, 87(3), 355–74.

[38] van der Velden, R.K.W. and M.S.M. van Smoorenburg (1997), "The Measurement of Overeducation and Undereducation: Self-report vs. Job Analys Method", Research Centre for Education and the Labour Market, Maastricht.

[39] Verdugo, R. and N.T. Verdugo (1989), "The Impact of Surplus Schooling on Earnings: Some Additional Findings", Journal of Human Resources 24(4), 629-695.

Tables and Figures

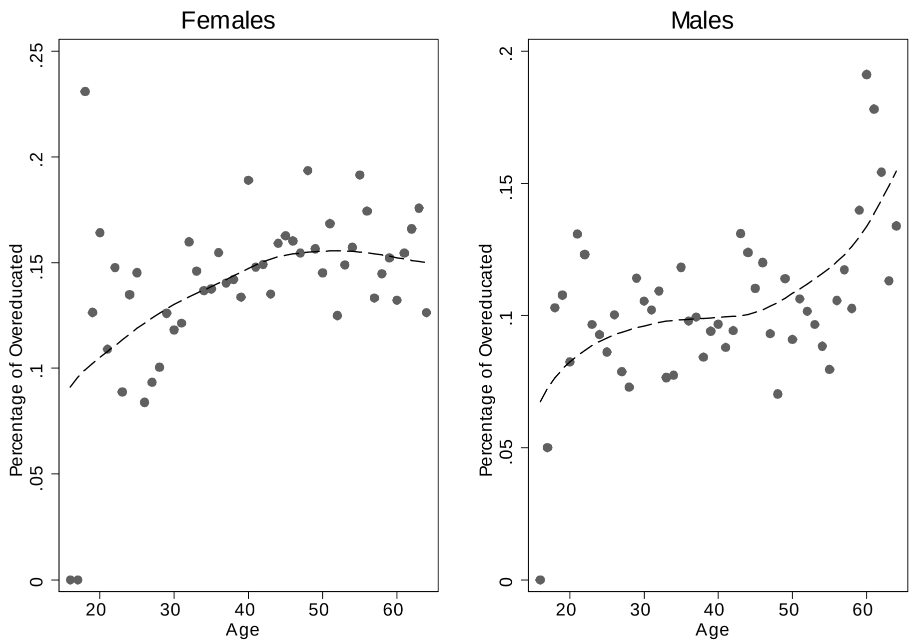

Figure 1: Incidence of overeducation by Gender and Age

.

Table 1: Summary statistics

	Well-Matched		*Mismatched*		H_0: *Equal Means*	
	Mean	s.d.	Mean	s.d.	t-value	p-value
Hourly wage	4.199	0.616	3.946	0.582	24.8	0.00
Male	0.480	0.500	0.386	0.487	11.4	0.00
Estonian origin	0.733	0.442	0.522	0.499	28.01	0.00
Married	0.686	0.464	0.654	0.476	4.0	0.00
Divorce/widowed	0.129	0.335	0.191	0.393	-10.8	0.00
Years of education	12.821	2.222	13.490	1.801	-18.4	0.00
Age	41.166	11.665	42.419	11.572	-6.4	0.00
Tenure	7.335	8.696	4.540	5.939	19.9	0.00
Public sector	0.333	0.471	0.273	0.446	7.6	0.00
Firm size 20-99	0.375	0.484	0.368	0.482	0.9	0.35
Firm size 100-499	0.170	0.375	0.176	0.381	-1.02	0.31
Firm size 500+	0.087	0.282	0.108	0.310	-4.38	0.00
Partime	0.075	0.263	0.116	0.320	-9.04	0.00
Agriculture	0.075	0.264	0.071	0.258	0.85	0.40
Fishing	0.005	0.071	0.004	0.060	1.23	0.22
Mining	0.015	0.123	0.021	0.144	-2.82	0.00
Manufacturing	0.250	0.433	0.311	0.463	-8.30	0.00
Elect, gas & water	0.027	0.162	0.022	0.148	1.76	0.08
Construction	0.064	0.245	0.045	0.208	4.67	0.00
Wholesale and retail	0.121	0.326	0.128	0.335	-1.33	0.18
Hotels and rest	0.022	0.148	0.032	0.176	-3.85	0.00
Transport	0.088	0.283	0.068	0.251	4.33	0.00
Financial interm	0.009	0.094	0.007	0.081	1.51	0.13
Real estate	0.042	0.202	0.064	0.246	-6.32	0.00
Public admin.	0.064	0.244	0.031	0.172	8.41	0.00
Education	0.114	0.318	0.097	0.296	3.26	0.00
Health	0.066	0.249	0.050	0.217	4.08	0.00
Others	0.036	0.187	0.049	0.216	-4.02	0.00

Note: Well-matched individuals: 29,288. Mismatched individuals: 4,332.

Table 2: The incidence of overeducation. Marginal and percentage effects from probit regressions

	Females			Males		
	meff	s.e.	pvalue	meff	s.e.	pvalue
Estonian origin	-0.013	0.002	0.000	-0.010	0.002	0.000
Public sector	-0.012	0.012	0.307	0.019	0.012	0.117
Partime	0.060	0.014	0.000	0.046	0.017	0.006
Married	0.010	0.010	0.312	-0.008	0.010	0.409
Divorced/widowed	0.035	0.012	0.003	0.041	0.017	0.015
Years of education	0.016	0.002	0.000	0.027	0.002	0.000
Aged 30-39	0.028	0.010	0.005	-0.001	0.009	0.901
Aged 40-49	0.066	0.011	0.000	0.014	0.010	0.147
Aged 50+	0.080	0.012	0.000	0.039	0.009	0.000
Tenure (2-3y]	-0.048	0.011	0.000	-0.040	0.008	0.000
Tenure (3-5y]	-0.101	0.011	0.000	-0.056	0.011	0.000
Tenure (5-10y]	-0.127	0.010	0.000	-0.091	0.009	0.000
Tenure (10y+	-0.154	0.008	0.000	-0.083	0.009	0.000
Firm size 20-99	0.005	0.008	0.582	0.007	0.008	0.328
Firm size 100-499	0.024	0.011	0.028	-0.006	0.008	0.447
Fishing	-0.135	0.047	0.004	-0.020	0.028	0.483
Mining	0.175	0.072	0.015	0.008	0.024	0.743
Manufacturing	-0.012	0.020	0.572	0.007	0.012	0.581
Elect, gas & water	0.039	0.045	0.386	-0.064	0.017	0.000
Construction	-0.077	0.034	0.022	-0.028	0.013	0.035
Wholesale and retail	-0.080	0.019	0.000	0.006	0.017	0.729
Hotels and rest	-0.033	0.025	0.197	0.008	0.031	0.791
Transport	-0.054	0.022	0.013	-0.048	0.014	0.000
Financial interm	-0.104	0.032	0.001	-0.098	0.023	0.000
Real estate	-0.017	0.024	0.466	0.000	0.021	0.997
Public admin.	-0.121	0.021	0.000	-0.092	0.014	0.000
Education	-0.082	0.022	0.000	-0.062	0.021	0.003
Health	-0.099	0.022	0.000	-0.019	0.030	0.517
Others	-0.013	0.027	0.616	-0.028	0.018	0.128

Note: Standard errors are robust and clustered at the individual level. * and ** denote significant at the 5 and 1 percent level respectively (two-tailed tests).

Table 3: Educational mismatch and wages in Estonia

M	Estimator	Females		Males	
		ATT	s.e.	ATT	s.e.
–	Mean difference	-0.271	0.012	-0.189	0.016
–	OLS	-0.278	0.011	-0.242	0.015
1	Simple matching	-0.241	0.014	-0.238	0.021
1	Bias-adjusted	-0.236	0.014	-0.246	0.021
4	Simple matching	-0.253	0.012	-0.224	0.017
4	Bias-adjusted	-0.242	0.012	-0.236	0.017

Note: Other covariates include language, marital status, years of education, age, age squared and time and regional dummies. Number of observations: Females (Treated=2,528, Controls=14,754, Total=17,282), Males (Treated=1,585, Controls=13,635, Total=15,220).

Table 4: Educational mismatch and wages by age categories

Age	**Estimator**	**Females**			**Males**		
		ATT	**s.e.**	(obs.)	**ATT**	**s.e.**	(obs.)
16-29							
	Mean Difference	-0.085	0.033	N_1=324	-0.109	0.035	N_1=354
	OLS	-0.077	0.032	N=2,766	-0.112	0.033	N=3,649
	M=1 Simple matching	-0.041	0.047		-0.080	0.039	
	M=1 Bias-adjusted	-0.019	0.046		-0.077	0.039	
	M=4 Simple matching	-0.060	0.039		-0.097	0.035	
	M=4 Bias-adjusted	-0.059	0.039		-0.136	0.035	
30-39							
	Mean Difference	-0.309	0.036	N_1=589	-0.192	0.037	N_1=344
	OLS	-0.273	0.027	N=4,224	-0.188	0.035	N=3,569
	M=1 Simple matching	-0.259	0.030		-0.245	0.046	
	M=1 Bias-adjusted	-0.239	0.029		-0.249	0.046	
	M=4 Simple matching	-0.261	0.024		-0.212	0.036	
	M=4 Bias-adjusted	-0.242	0.025		-0.215	0.036	
40-49							
	Mean Difference	0.273	0.019	N_1=845	-0.215	0.032	N_1=420
	OLS	-0.295	0.017	N=5,268	-0.293	0.031	N=4,009
	M=1 Simple matching	-0.256	0.021		-0.236	0.037	
	M=1 Bias-adjusted	-0.254	0.021		-0.261	0.037	
	M=4 Simple matching	-0.279	0.018		-0.247	0.029	
	M=4 Bias-adjusted	-0.270	0.018		-0.282	0.029	
50-64							
	Mean Difference	-0.320	0.028	N_1=770	-0.217	0.028	N_1=467
	OLS	-0.344	0.021	N=5,024	-0.344	0.026	N=3,993
	M=1 Simple matching	-0.295	0.029		-0.323	0.039	
	M=1 Bias-adjusted	-0.295	0.029		-0.354	0.039	
	M=4 Simple matching	-0.315	0.024		-0.306	0.030	
	M=4 Bias-adjusted	-0.306	0.024		-0.332	0.030	

Note: Other covariates include language, marital status, years of education and time and regional dummies.

Table 5: Robustness check. Including additional controls.

Age	Estimator	Females			Males		
		ATT	**s.e.**	(obs.)	**ATT**	**s.e**	(obs.)
ALL							
	Mean Difference	-0.271	0.012	N_1=2,528	-0.189	0.016	N_1=1,585
	OLS	-0.252	0.011	N=17,282	-0.205	0.015	N=15,220
	Matching	-0.232	0.015		-0.213	0.021	
16-29							
	Mean Difference	-0.085	0.033	N_1=324	-0.109	0.035	N_1=354
	OLS	-0.072	0.032	N=2,766	-0.089	0.033	N=3,649
	Matching	-0.063	0.051		-0.068	0.040	
30-39							
	Mean Difference	-0.309	0.036	N_1=589	-0.192	0.037	N_1=344
	OLS	-0.263	0.027	N=4,224	-0.148	0.035	N=3,569
	Matching	-0.248	0.033		-0.172	0.046	
40-49							
	Mean Difference	0.273	0.019	N_1=845	-0.215	0.032	N_1=420
	OLS	-0.287	0.017	N=5,268	-0.256	0.030	N=4,009
	Matching	-0.307	0.023		-0.287	0.037	
50-64							
	Mean Difference	-0.320	0.028	N_1=770	-0.217	0.028	N_1=467
	OLS	-0.282	0.021	N=5,024	-0.296	0.025	N=3,993
	Matching	-0.243	0.030		-0.258	0.034	

Note: Other covariates include language, marital status, years of education, tenure, tenure squared, public sector dummy, sectoral dummies, time and regional dummies.

Table 6: Matching quality: mean differences in covariates pre and post matching

	Before Match			After Match		
Variable	**Wellmatch.**	**Overedu.**	**Diff.**	**Wellmatch.**	**Overedu.**	**Diff.**
			Females			
Estonian origin	0.036	-0.209	0.245	-0.075	-0.101	0.026
Married	0.006	-0.034	0.040	0.000	-0.018	0.018
Divorced/widowed	-0.042	0.247	-0.289	0.084	0.089	-0.004
Years of education	-0.003	0.020	-0.024	0.133	0.124	0.008
Age	-0.004	0.021	-0.025	0.044	0.059	-0.015
County dummy 1	-0.068	0.396	-0.464	0.223	0.221	0.002
County dummy 2	-0.058	0.341	-0.400	0.057	0.057	0.000
County dummy 3	0.045	-0.263	0.308	-0.053	-0.053	0.000
County dummy 4	-0.088	0.515	-0.604	0.232	0.232	0.000
County dummy 5	0.095	-0.553	0.647	-0.109	-0.109	0.000
County dummy 6	-0.039	0.230	-0.269	0.049	0.049	0.000
County dummy 7	0.015	-0.089	0.105	-0.024	-0.024	0.000
County dummy 8	0.024	-0.143	0.167	-0.045	-0.045	0.000
			Males			
Estonian origin	0.018	-0.154	0.172	-0.052	-0.074	0.022
Married	0.003	-0.024	0.026	-0.035	-0.050	0.015
Divorced/widowed	-0.059	0.506	-0.565	0.057	0.093	-0.036
Years of education	-0.009	0.073	-0.082	0.340	0.392	-0.052
Age	-0.003	0.026	-0.029	-0.011	0.026	-0.037
County dummy 1	-0.046	0.395	-0.441	0.224	0.223	0.001
County dummy 2	-0.048	0.411	-0.459	0.082	0.082	0.000
County dummy 3	0.014	-0.116	0.130	-0.026	-0.026	0.000
County dummy 4	-0.040	0.345	-0.385	0.195	0.195	0.000
County dummy 5	0.023	-0.195	0.218	-0.038	-0.038	0.000
County dummy 6	0.009	-0.075	0.084	-0.032	-0.032	0.000
County dummy 7	-0.041	0.352	-0.393	0.033	0.033	0.000
County dummy 8	0.036	-0.307	0.343	-0.095	-0.095	0.000

Note: Simple matching bias adjusted (specification contained in Table 3). County dummies 1 to 8 refer to, respectively: Tallinn, Harju (excl. Tallinn), Hiiu, Ida-Viru, Jogeva, Jarva, Laane, Laane-Viru. Additional county dummies (not shown for clarity) presented identical results. These are: Polva, Parnu, Rapla, Saare, Tartu, Valga, Viljandi and Voru.

Table 7: Sensitivity analysis: effect of 'calibrated' confounders

	Females					
	Fraction U=1					
	$\mathbf{p}_{11}$	$\mathbf{p}_{10}$	$\mathbf{p}_{01}$	$\mathbf{p}_{00}$	**ATT**	**s.e.**
No confounder	0	0	0	0	-0.236	0.014
Neutral confounder	0.50	0.50	0.50	0.50	-0.240	0.016
Confounder distributed like age	0.54	0.42	0.48	0.45	-0.239	0.017
Other confounders:						
a)	0.90	0.80	0.20	0.10	-0.159	0.024
b)	0.90	0.70	0.20	0.10	-0.162	0.023
c)	0.90	0.60	0.20	0.10	-0.165	0.023
d)	0.90	0.50	0.20	0.10	-0.167	0.023
e)	0.90	0.50	0.30	0.10	-0.130	0.021
f)	0.90	0.50	0.40	0.10	-0.104	0.019
g)	0.90	0.50	0.40	0.00	-0.031	0.017

Note: The first four columns contain the parameters p_{ij} used to simulate the binary confounder (U) in the way described in section 6. The subsequent columns contain the simulated ATT when controlling for U over 100 iterations and their standard error (s.e.).

Table 8: Sensitivity analysis: effect of 'calibrated' confounders

	Males					
	Fraction U=1					
	$\mathbf{p}_{11}$	$\mathbf{p}_{10}$	$\mathbf{p}_{01}$	$\mathbf{p}_{00}$	**ATT**	**s.e.**
No confounder	0	0	0	0	-0.246	0.021
Neutral confounder	0.50	0.50	0.50	0.50	-0.243	0.022
Confounder distributed like age	0.53	0.49	0.49	0.45	-0.241	0.023
Other confounders:						
a)	0.90	0.80	0.20	0.10	-0.142	0.030
b)	0.90	0.70	0.20	0.10	-0.153	0.030
c)	0.90	0.60	0.20	0.10	-0.158	0.029
d)	0.90	0.50	0.20	0.10	-0.162	0.029
e)	0.90	0.50	0.30	0.10	-0.118	0.027
f)	0.90	0.50	0.40	0.10	-0.092	0.025
g)	0.90	0.50	0.40	0.00	0.052	0.024

Note: The first four columns contain the parameters p_{ij} used to simulate the binary confounder (U) in the way described in section 6. The subsequent columns contain the simulated ATT when controlling for U over 100 iterations and their standard error (s.e.).