College degree supply, productivity spillovers and occupational allocation of graduates in Central European countries*

Barbara Gebicka
CERGE-EI

Anna Lovasz
IE HAS

June 2010

Abstract

Public funding drives much of the recent growth of college degree supply in Europe, but few indicators are available to assess its optimal level. In this paper, we investigate an indicator of college skills usage – the fraction of college graduates employed in "college" occupations. Gottschalk and Hansen (2003) propose to identify "college" occupations based on within-occupation college wage premia; we build on their strategy to study the local-labor-market relationship between the share of college graduates in the population and the use of college skills. Empirical results based on worker-level data from NUTS-4 districts in the Czech Republic, Hungary and Slovakia suggest a positive relationship, thus supporting the presence of an endogenous influence of the number of skilled workers on the demand for them.

*This research has been supported by the IX Regional Research Competition GDN grant No. RRC IX-51.

1 Introduction

While primary and some form of secondary education is available for the vast majority of citizens in the developed countries, higher education is only accessible to a limited number of people. These limits are partially driven by public funds devoted to higher education, which is especially binding in countries where the majority of higher education institutions are public. As both the over- and undersupply of college seats could result in efficiency losses for society, there is a need to understand the forces shaping the demand for skilled labor to inform policy decisions concerning the provision of higher education.

Recent economic literature has approached the topic of optimal level of college degree supply by analyzing different indicators of college skills utilization. The most straightforward is to analyze social returns[1] to higher education (Acemoglu and Angrist 2000, Moretti 2004), which directly capture the benefits of educating people, however, are difficult to measure. An alternative is offered by the overskilling literature (see McGuiness 2006 for a review), which investigates employment of college graduates in the so-called "noncollege" occupations (Pryor and Schaffer 1997, McGuiness and Bennett 2007) in order to quantify the oversupply of college skills. This line of research offers an easy to measure indicator of college skills usage which is not, however, supported by an economic model. Only recently, Gottschalk and Hansen (2003) proposed a methodology for classifying occupations into "college" and "noncollege" based on a rigorous, though simple, model. This equips us with a more reliable tool to measure the fraction of college graduates employed in "noncollege" occupations – an indicator useful in assessing whether changes in the supply of skilled labor meet changes in the demand for them. In this paper, we use the measure of college graduates employed in "noncollege" occupations, as proposed by Gottschalk and Hansen (2003), to find out whether an increased number of college graduates attracts firms using advanced technologies and thus triggers a shift in the demand for skilled labor.

The model proposed by Gottschalk and Hansen (2003) assumes that "noncollege" oc-

[1] There is also a vast stream of literature on private returns to higher education, known as the college wage premium, and their connection to the relative supply and demand for skilled labor (Bound and Johnson 1992, Katz and Autor 1999, Fortin 2006). As the college wage premium is a relative measure of returns to higher education, it is not informative of the absolute demand for college graduates.

cupations do not value college-gained skills and thus pay none or very little wage premium to college graduates, while “college” occupations pay a significant college wage premium. This property allows us to order occupations according to their estimated returns to college and to classify as “college” those occupations which fall above a certain threshold. Several studies follow this approach to measure the fraction of college graduates employed in “noncollege” occupations in the U.S. (Gottschalk and Hansen, 2003), Portugal (Cardoso, 2007), and the U.K. (Grazier, 2008).[2] These papers only analyze the time trend of the overskilling measure at the aggregate level. It would be more informative, however, to see whether the extent of overskilling is correlated with the number of college graduates in the economy. This relationship is depicted in Figure 1, which plots the probability of a young college graduate to be employed in a “noncollege” occupation,[3] as reported by the authors of the above-mentioned articles, against the fraction of college graduates in the young population.[4] This figure also presents an analogous relationship for the Czech Republic and Hungary, the countries analyzed in more detail in this paper.

Two features stand out in Figure 1. First, within a country the probability of a college graduate to work in a “noncollege” occupation is negatively correlated with the fraction of college graduates in the population. Second, in countries with a higher proportion of highly educated people in the population, the likelihood of observing a college graduate work in a “noncollege” occupation is higher. The latter observation could be an artifact of the constant college wage premium threshold used in these studies to distinguish between “college” and “noncollege” occupations. It is generally understood that economies with a relatively low endowment of skilled labor report high college premia (Brunello et. al 2000, Card and Lemieux 2001), which could be reflected in more occupations being classified as “college” in these countries. More robust and more interesting is the positive within-country correlation between the fraction of college graduates in the population and the probability of

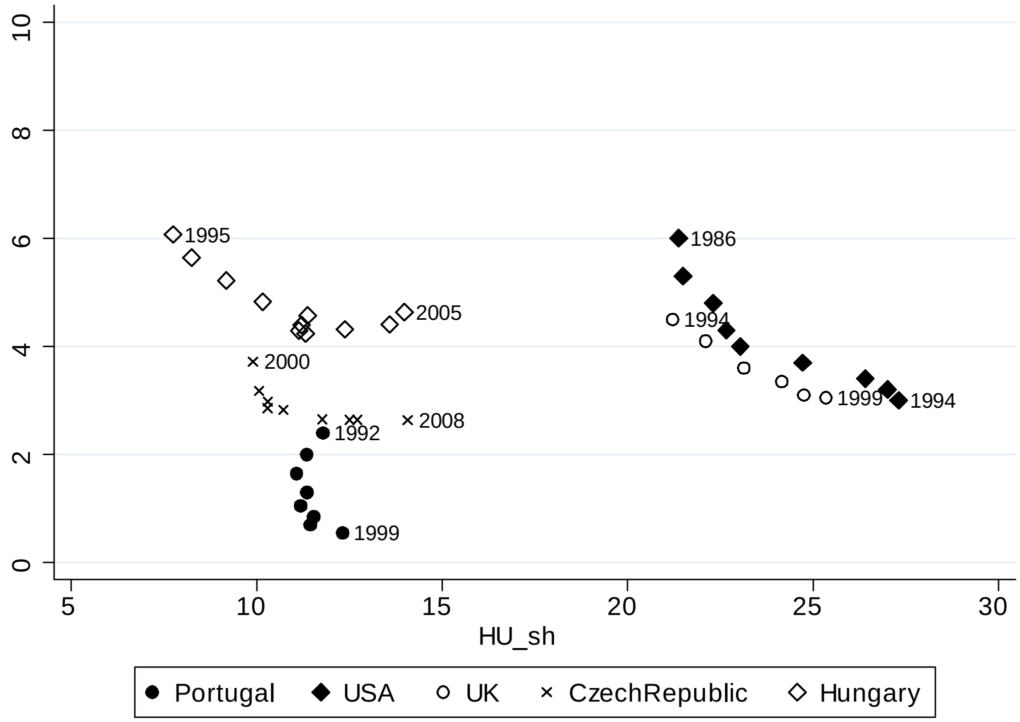
[2] Galasi (2005) analyzed the employment of college graduates in “noncollege” occupations in Hungary. His results, although qualitatively the same as those of other authors, are not comparable, as he uses different threshold to distinguish between “college” and “noncollege” occupations.

[3] The probability of being employed in a "noncollege" occupation is a disaggregated measure of the fraction of college graduates employed in “noncollege” occupations.

[4] A young population is defined as 20-39 years of age.

a college graduate to work in a "noncollege" occupation. Following a simple supply-demand analysis, one would expect the opposite relationship.[5] Thus, it is tempting to interpret this feature as the positive influence of an increased number of skilled workers on the number of skill-intensive positions offered by firms (i.e. as a spillover effect). Yet, the observed correlation could be spurious and reflect just the simultaneous reaction of the demand and supply side of the labor market for college graduates to positive technological shocks.

Figure 1: Propensity of a college graduate to work in a "noncollege" occupation vs. the share of college graduates in the labor force across countries.



Source: Authors' compilation using Gottschalk and Hansen (2003), Cardoso (2007), Grazier et al. (2008), Eurostat, and U.S. Census Bureau as well as the ISAE and WES data.

To better understand the patterns observed in Figure 1, we extend the Gottschalk and Hansen (2003) setup to explicitly model the relationship between the number of college graduates available in the labor market and the fraction of them working in "noncollege"

[5] This is a consequence of movement along a downward-sloping demand curve.

occupations. Instead of working with an aggregate time trend, we estimate this relationship using the within country cross-regional variation in the fraction of college graduates working in "noncollege" occupations. This approach not only allows us to use more data points but also makes it easier to break the simultaneity between the number of college graduates in the market and their occupational allocation. Cross-regional patterns are similar to those observed within a country over time. The relationship of interest is found to be negative in the cross-regional analysis which suggests that the long-run equilibrium is shaped by the endogenous influence of the number of skilled workers on the demand for them.

The analysis presented in this paper concentrates on the Czech Republic and Hungary. These Central European countries are especially interesting because their higher education systems have been expanding rapidly but unequally in recent years, resulting in significant between-year and across-region variation in the educational structure of the population.[6] Moreover, as Central European countries are still lagging behind the Western economies in terms of technological development, there is a lot of opportunity for technological progress to happen and advanced capital to flow in. Finally, the choice of these countries adds policy relevance to this research. The higher education systems in these countries is largely state-funded and thus the provision of college education is a public policy decision. Awareness of the channels which affect the demand for college-educated labor would facilitate decision-making concerning the extent of higher education expansion. In the absence of the endogenous effect college enrolments should simply reflect the trend in technological progress of the economy; while the existence of this effect implies that increasing the educational attainment of the local population could be used to attract advanced technologies and to increase the skill bias of the economy.

The reminder of the paper is organized as follows. Section 2 places this study in the context of the existing literature. The theoretical and empirical models of college and high school graduates' allocation across different occupations are described in Sections 3 and 4, respectively, followed by a definition of "college "and "noncollege" occupations in Section 5. Estimation of the causal relationship between the relative stock of college graduates and the

[6] More about the higher education expansion in the Czech Republic could be found in Gebicka (2010) and in Hungary in Lannert (2005).

fraction of them working in "noncollege" occupations is then discussed. Section 7 concludes.

2 Demand for College Graduates in the Literature

Several streams of literature are related to this paper. First, Acemoglu (2002, 2003) suggests that the extent of the skill bias of technology, and thus the demand for skills, can be shifted endogenously by intense international trade and by the presence of many skilled workers. Similar conclusions are reached by Moretti (2004), who shows that a high concentration of college-educated workers in a city's population has a positive effect on wages of all education groups in that city, including the college graduates. This implies the existence of positive productivity spillovers from the spatial concentration of skills and suggests that a large number of college graduates in a labor market can trigger a shift in the demand for them. Fortin's (2006) findings of a negative relationship between the production of college graduates and the college-high school wage gap across the U.S. states suggest that the positive effect of a high concentration of college graduates on local wages is stronger for high school-educated workers. These findings are challenged by Bound et. al (2004), who find that the production of college graduates in U.S. states does not correspond to their stock, because of a significant level of migration. If this is also true for the countries analyzed in this study, its policy implications could be limited. Nevertheless, it is generally known that in Central Europe both the within-country and across-countries mobility of labor is much lower than in the U.S. (e.g. Fidrmuc 2004) and enrolments in higher institutions translate into a future supply of college graduates to local labor markets in these countries. Thus, to identify potential endogenous shifts in the demand for labor, we follow Moretti (2004) and investigate the relationship between the presence of college-educated individuals in the economy and the demand for skilled labor. However, instead of analyzing college graduates' wages, we investigate their occupational allocation as the indicator of college skills usage.

Occupational allocation of college graduates is the central focus of another stream of literature related to this paper, widely known as the overeducation (overskilling) literature. Studies in this field measure the fraction of college graduates employed in occupations not requiring a college degree and estimate the wage effects of being employed in such an occupation. They find that the incidence of overeducation is increasing over time (Walker and Zhu 2005, evidence for the U.K.) and that it is associated with significant wage punishment (McGuiness 2006, a metastudy) which, however, is largely reduced if individual heterogeneity is taken into account (Bauer 2002, evidence for Germany). This literature typically classifies individuals as being overskilled if they work in an occupation which has the median (average) year of schooling lower than that of the individual, has the official schooling requirement, as defined in the job description, lower than that of the individual or is assessed by the individual to require lower skills than she has. While this line of research studies a phenomenon directly reflecting the demand for college graduates, it suffers from the lack of an economic model supporting the measures of overskilling. This gap is filled by Gottschalk and Hansen (2003), who develop a simple supply-demand framework which models the allocation of college graduates between "college" and "noncollege" occupations. We depart from their model when investigating the occupational allocation of college graduates.

Research on the demand for college-educated workers has not been that extensive in the context of Central Europe. The only comparative study by Flabbi et al. (2008) shows that the returns to education were increasing or stayed constant in several Central and Eastern European countries throughout transition. Analyses concentrating on the Czech and Slovak Republics, e.g. Filer et al. (1999) Jurajda (2005), Munich et al. (2005), and Hungary (Kertesi and Kollo 2005, Galasi 2007) also confirm this finding. Another study by Jurajda (2004) shows that college graduates' wages are insensitive to their concentration across Czech districts. In a related work, Jurajda and Terrell (2007) find that significant differences in unemployment rates across regions of post-communist economies can be to a large extent explained by variations in local human capital endowment. Additionally, they show that FDI flows to regions characterized by higher human capital endowment, which is in line with Acemoglu's hypothesis of endogenous technological progress. Our study falls into this line of research, as it investigates the relationship between the educational structure of the local population and the labor market situation of college graduates.

3 Theoretical framework

In this paper we analyze the influence of variations in the relative number of college graduates in the population on their allocation between "college" and "noncollege" occupations. The first question to be answered before proceeding to the empirical analysis is why we would observe some college graduates working in "noncollege" occupations, and how to recognize which occupations are "college" and which are "noncollege". A model dealing with these issues has been proposed by Gottschalk and Hansen (2003). We modify it to directly model the influence of supply and demand conditions on the equilibrium allocation of college graduates. Later on, we also allow for endogenous influence of the number of college graduates in the labor market on their productivity in "college" occupations. This leads to an ambiguous prediction of the sign of the relationship between the relative number of college graduates in the population and their occupational allocation.

The model proposed by Gottschalk and Hansen (2003) assumes that there are two sectors in the economy: a "college" sector and a "noncollege" sector. Competitive firms in both sectors produce the same uniform good.[7] They have the following production functions:

$$Q_1 = F_1(\alpha_{C1}L_{C1} + \alpha_{N1}L_{N1}) \tag{1}$$

$$Q_2 = F_2(\alpha_{C2}L_{C2} + \alpha_{N2}L_{N2}), \tag{2}$$

where Q_j measures the output of sector j, L_{Cj} and L_{Nj} are the amounts of college- and high school-educated labor in sector j, α_{ij} are productivities of labor type i in sector j, and F_js are twice-differentiable functions with $F_j'(\cdot) > 0$ and $F_j''(\cdot) < 0$. It is assumed that in sector 1 college-educated labor is relatively more productive than high school-educated labor as compared to sector 2 ($\frac{\alpha_{C1}}{\alpha_{N1}} > \frac{\alpha_{C2}}{\alpha_{N2}}$). That is why sector 1 is called the "college" sector.

Firms' profit maximization under the price of output normalized to unity and labor input prices being w_{C1}, w_{C2}, $w_{N1}$and w_{N2}, respectively, gives the following condition:

$$\frac{w_{C1}}{w_{N1}} = \frac{\alpha_{C1}}{\alpha_{N1}} > \frac{\alpha_{C2}}{\alpha_{N2}} = \frac{w_{C2}}{w_{N2}}, \tag{3}$$

i.e. the wages of college graduates relative to high school graduates are higher in sector 1,

[7] Allowing the two sectors to produce different goods does not influence the inference of this model. This assumption is kept for the purpose of clarity.

the "college" sector. This property will be further used to distinguish between "college" and "noncollege" occupations.

To complete the model, we modify the supply functions of different labor types to both sectors proposed by Gottschalk and Hansen (2003). Like these authors, we assume that workers in a pool of all college and high school graduates decide to work in either sector "*based on their heterogenous preferences and the relative wages available to them across sectors*" (p. 5). On top of that, however, we specify the relationship between the total number of college and high school graduates in the labor market and the sector-specific supply functions, which is not explicitly shown in the original model.[8] The authors do not need to model this because they do not analyze the relationship between the structure of the labor force and the allocation of workers across occupations. In our version of the model it is assumed that the total supply of a given labor type to a given sector is a proportion of all workers of this type in the population. This allows for direct analysis of the influence of changes in the structure of the labor force on the market equilibrium. The assumed supply functions are the following:

$$\ln\left(\frac{L_{C1}^S}{L_C}\right) = \lambda_C + \beta_C \ln\left(\frac{w_{C1}}{w_{C2}}\right) \tag{4}$$

$$L_{C2}^S = L_C - L_{C1}^S \tag{5}$$

$$\ln\left(\frac{L_{N1}^S}{L_N}\right) = \lambda_N + \beta_N \ln\left(\frac{w_{N1}}{w_{N2}}\right) \tag{6}$$

$$L_{N2}^S = L_N - L_{N1}^S, \tag{7}$$

where L_C and L_N are the total numbers of college and high school graduates in the labor market, and β_i and λ_i are the aggregate preference parameters of workers of type i.

Together, equations (3)[9] and (4) - (7) define the equilibrium allocation and wages of college and high school graduates among the two sectors. An important property of this model

[8]The supply functions of college and high school graduates to the "college" sector used by Gottschalk and Hansen (2003) are the following: $L_{C1}^S = \lambda_C + \beta_C \frac{w_{C1}}{w_{C2}}$ and $L_{N1}^S = \lambda_N + \beta_N \frac{w_{N1}}{w_{N2}}$. Note that they do not explicitly account for the total amount of college- and high school-educated labor in the economy.

[9]Equation (3) actually consists of 4 equations: $w_{C1} = \alpha_{C1}F_1'(L_1)$, $w_{N1} = \alpha_{N1}F_1'(L_1)$, $w_{C2} = \alpha_{C2}F_2'(L_2)$, and $w_{N2} = \alpha_{N2}F_2'(L_2)$, where $L_1 = \alpha_{C1}L_{C1} + \alpha_{N1}L_{N1}$ is the total labor aggregate used in sector 1 and $L_2 = \alpha_{C2}L_{C2} + \alpha_{N2}L_{N2}$ is the total labor aggregate used in sector 2.

is that in equilibrium there are some college-educated workers employed in both sectors. This study concentrates on the fraction of college graduates working in the "noncollege" sector, which is defined as

$$\pi_C \equiv \frac{L_{C2}}{L_C}. \tag{8}$$

The main advantage of the proposed model is that it directly captures the influence of the supply conditions (the total amount of each labor type in the economy, L_i) and demand conditions (labor productivities, α_{ij}) on the equilibrium fraction of college graduates working in the "noncollege" sector (π_C^*).

$$\pi_C^* \equiv 1 - \frac{L_{C1}^*}{L_C} = f\left(L_C, L_N, \alpha_{C1}, \alpha_{N1}, \alpha_{C2}, \alpha_{N2}\right). \tag{9}$$

To understand the forces influencing the occupational allocation of college graduates, let us analyze how the equilibrium fraction of college graduates working in the "noncollege" sector reacts to the shifts in supply- and demand-characterizing variables, i.e. the structure of the labor market ($\frac{L_C}{L_N+L_C}$) and the extent of the skill bias of technology ($\frac{\alpha_{C1}}{\alpha_{N1}}$).

First, we analyze how the equilibrium allocation changes when the skill-biased technological change (SBTC) happens in the "college" sector, i.e., when $\frac{\alpha_{C1}}{\alpha_{N1}}$ grows and all other variables are kept unchanged. This change should increase wages offered by firms in the "college" sector to college graduates (demand for college graduates in sector 1 shifts up). Higher wages attract more college graduates to the "college" sector, as described by equation (4). This, in turn, lowers a bit their wages in sector 1 and increases their wages in sector 2. Finally, wages adjust in such a way that no more workers want to change jobs. The new equilibrium is characterized by higher wages for college graduates in both sectors, but wages in sector 1 increase more as compared to the initial level. This makes the new $\frac{w_{C1}^*}{w_{C2}^*}$ higher than the initial one and thus the new π_C^* lower than the initial one. To sum up,

$$\frac{\partial \pi_C^*}{\partial\left(\alpha_{C1}/\alpha_{N1}\right)} < 0. \tag{10}$$

Next, let us analyze what happens when the relative stock of college graduates in the labor market ($\frac{L_C}{L_N+L_C}$) increases, which is a result of growth in L_C and a related fall in L_N. This change results in an upward shift in the supply of college graduates and a downward shift in the supply of high school graduates to both sectors, as shown by equations (4) and

(6). As a result, wages of all labor types in the "college" sector fall. In the "noncollege" sector wages fall as well, but less dramatically, as long as $\frac{\alpha_{C2}}{\alpha_{N2}} \geqslant 1$. If $\frac{\alpha_{C2}}{\alpha_{N2}} < 1$, wages in sector 2 may actually rise. In any case, the ratio $\frac{w_{C1}}{w_{C2}}$ falls and some workers reallocate from the "college" to the "noncollege" sector. This, in turn, lowers a bit wages in sector 2 and increases them in sector 1 (but not above the initial level) so that ultimately nobody wants to change jobs. The new equilibrium is characterized by lower wages for college graduates in both sectors, but wages in sector 1 decrease more as compared to the initial level. This makes the new $\frac{w^*_{C1}}{w^*_{C2}}$ lower than the initial one and thus the new π^*_C higher than the initial one. To sum up,

$$\frac{\partial \pi^*_C}{\partial \left(\frac{L_C}{L_N+L_C}\right)} > 0. \tag{11}$$

The above analysis leads to the following formulation of the relationship between the relative supply of college graduates to the labor market and the fraction of them working in "noncollege" occupations:

$$\pi^*_C = f\left(\underset{+}{\frac{L_C}{L_N+L_C}}, \underset{-}{\frac{\alpha_{C1}}{\alpha_{N1}}}, \textit{other factors}\right). \tag{12}$$

Assuming that the relationship is approximately linear[10] and other factors vary randomly, it can be written it in the following form:

$$\pi^*_C = \gamma_0 + \gamma_1 \frac{L_C}{L_N+L_C} + \gamma_2 \frac{\alpha_{C1}}{\alpha_{N1}} + \varepsilon, \tag{13}$$

where $\gamma_1 > 0$ and $\gamma_2 < 0$, as derived.

According to the model presented above, the relationship between $\frac{L_C}{L_N+L_C}$ and π^*_C is positive. However, this model does not take into account the endogenous influence of the labor force structure on college graduates' productivity in "college" occupations. Let us now introduce endogeneity (also known as productivity spillover) into the model to show that it can alter the relationship. A general representation of productivity spillovers commonly used in the literature is in the form of productivity being an increasing function of aggregate skills (e.g., Acemoglu and Angrist 2000, Moretti 2004). In this paper we use a simple linear

[10] The model outlined in this section has no closed form solution. Therefore, we have to approximate its functional form.

relationship:

$$\frac{\alpha_{C1}}{\alpha_{N1}} = \alpha + \delta \frac{L_C}{L_N + L_C}, \tag{14}$$

where $\delta \geq 0$ ($\delta = 0$ implies no spillovers and $\delta > 0$ implies the existence of positive productivity spillovers). Incorporating this into equation (13), we get:

$$\pi_C^* = \gamma_0 + \underbrace{(\gamma_1 + \gamma_2 \delta)}_{\theta} \frac{L_C}{L_N + L_C} + \gamma_2 \cdot \alpha + \varepsilon. \tag{15}$$

When allowing for productivity spillovers from a high concentration of skills, the sign of the relationship between the relative supply of college graduates and the fraction of them working in "noncollege" occupations is not clearly predicted by the model. If the direct effect (γ_1) is stronger than the spillover effect ($\gamma_2\delta$), the overall relationship is negative; however, if the spillover effect is strong enough to compensate for the direct effect, the overall relationship is positive. The goal of this paper is to estimate the parameter $\theta_1 \equiv \gamma_1 + \gamma_2\delta$ to determine whether positive or negative effects prevail in the influence of the relative stock of college graduates on their allocation across occupations.

Before proceeding to the empirical analysis, let us discuss the assumptions behind the model and the limitations implied by them. First of all, it is important to acknowledge that the above model describes a single closed economy. One should be careful when applying it to compare districts within one country if workers and firms are mobile. In the context of the Central European countries, however, mobility of labor is limited. As shown by or Fidrmuc (2004) for the Visegrad countries or Cseres-Gergely (2002) for Hungary and Gebicka (2010) for the Czech Republic, workers tend to stay in the district where they graduated. Additionally, there are other factors than labor availability influencing firms' decisions to locate in a given district, and thus firm mobility does not fully compensate cross-district differences in the labor force structure. This allows us to treat districts as separate labor markets and use equation (15) to analyze the cross-district relationship between the relative supply of labor and the fraction of college graduates working in "noncollege" occupations.

Second, the assumption of workers' heterogeneous preferences towards job attributes could be questioned. While this is the only approach used in this line of literature, one could come up with alternative explanations for why we observe college graduates in both "college" and "noncollege" occupations. Workers might have heterogeneous ability to use college-gained skills, and "college" firms employ only those with high enough ability. Alternatively, the amount of capital complementing college-educated workers might be limited, which sanctions the number of college graduates who can be employed in "college" occupations. Discussion of these models is not within the scope of this paper. Let us note, however, that each of the alternative explanations supports the prediction of the model used to classify occupations, i.e. that relative wages of college to high school graduates are higher in the "college" sector (see the earlier version of this paper: Gebicka 2010). We base the analysis on the Gottschalk and Hansen (2003) model to be consistent with the literature.

4 Estimation Strategy

The theoretical model derived in the previous section serves as a baseline for analyzing the relationship between the relative stock of college graduates and the fraction of them working in "noncollege" occupations. Before formulating an econometric model based on these derivations, let us note that equation (15) accommodates an implicit assumption that the aggregate preference of workers, summarized by parameters β_C, β_N and λ_C, λ_N, are constant within and across districts. This is, however, a very unrealistic assumption. It can be argued that the composition of characteristics of individuals living in a given district influences their allocation across occupations through their preference parameters. If, for example, in a given district there are many females with a college education (who are, on average, less flexible in looking for employment), there might be a higher fraction of college graduates in "noncollege" occupations there. In order to account for such effects, we formulate an econometric model on the individual rather than on the aggregate level, i.e., we model the propensity of an individual college graduate to work in a "noncollege" occupation as a function of her characteristics and characteristics of the region where she lives, as shown in equation (16). This model can be thought of as a disaggregated version of equation (15).

$$\text{Prob}(nocollege_{ikt}) = \gamma_0 + \mathbf{X}'_{ikt}\boldsymbol{\theta}_0 + \theta_1 \left(\frac{L_C}{L_N + L_C} \right)_{kt} + \mathbf{Y}'_{kt}\boldsymbol{\theta}_2 + \varepsilon_{ikt}, \tag{16}$$

where $\text{Prob}(nocollege_{ikt})$ is an indicator whether a college graduate i in district k at time t is working in a "noncollege" occupation, $\mathbf{X}'_{ikt}$ is a vector of individual characteristics such as the worker's potential labor market experience (in years) and gender, $\left(\frac{L_C}{L_N+L_C}\right)_{kt}$ is the relative stock of college graduates in district k at time t, $\mathbf{Y}'_{kt}$ is a vector of other year-district specific characteristics, and ε_{ikt} represents the individual, time and district specific unobservable determinants of college graduates' allocation across occupations. The parameter of main interest is θ_1; it describes the causal relationship between the relative number of college graduates in a district's population and their fraction working in "noncollege" occupations.[11]

The district specific characteristics in $\mathbf{Y}_{kt}$ include size measures such as the density of the district's population, and the logarithm of the district's labor force to account for assortative matching effects. It is generally accepted that in larger markets, workers and firms find each other more easily (Wheeler, 2001) and thus we could observe a lower fraction of college graduates working in "noncollege" occupations in large labor markets. We also control for the share of employment in the public sector because the individual level data used for estimations covers only employees from the commercial sector, while the public sector usually employs many college graduates, which can influence the district's equilibrium share of the highly educated.[12]

The source of identification used to estimate θ_1 is the variation in the fraction of highly educated adults across the NUTS-4 district populations and the simultaneous variation in the proportion of college graduates working in "noncollege" occupations in these districts. The identification is, however, complicated by the omitted variable problem. Some of the factors captured by the error term might bias the estimate of $\widehat{\theta}_1$ due to a correlation with the relative supply of college graduates. The major source of bias is the unobserved heterogeneity across districts, as well as over time, in the demand for labor.[13] Both time and

[11] Ideally, the above should be modeled as a choice between three alternatives: working in the "college" sector, working in the "noncollege" sector, and being unemployed. Unfortunately, the data set used in this paper does not contain information about the unemployed. Nevertheless, this is not an important issue in the case of the Czech Republic, where the unemployment rate of college graduates did not exceed 4.6% in any district over the 2000-2006 period.

[12] We have also experimented with using real GDP per capita as an additional explanatory variable, but it appears to have no power in explaining the variation in the fraction of college graduates working in "noncollege" occupations.

[13] An omitted variable bias might also result from workers sorting into districts according to their unobservable abilities. This could be addressed by controlling for workers' fixed effects. The data used in this study do have a repeated cross-section structure, which does not allow for this approach. Nevertheless, Moretti (2004) shows that omitted "*individual characteristics are not a major source*

district specific productivity shocks might partially drive the variation in the stock of college graduates. For example, the expansion of hi-tech industry in one district may attract highly educated workers to move there. In this case we expect $cov(\varepsilon_{kt}, \frac{L_C}{L_N+L_C}_{kt}) > 0$ (i.e., positive productivity shocks induce a higher fraction of college graduates), thus the OLS estimates of the relationship from equation (16) would be biased downwards.[14] As the supply of college places in the Central European countries was to a great extent determined by public authorities, also another mechanism creating the correlation between the productivity shocks and district-specific supply of college graduates is possible. If the government decides to expand higher education in disadvantaged districts, we expect $cov(\varepsilon_{kt}, \frac{L_C}{L_N+L_C}_{kt}) < 0$ (i.e., negative productivity shocks induce a higher fraction of college graduates) and the OLS estimates of the relationship from equation (16) would be biased upwards.

Endogeneity of the fraction of the population with a college degree can be overcome in several ways. The first proposal is to use an instrument that predicts well the share of college graduates in a district's population but at the same time is uncorrelated with district specific productivity shocks. In the search for an instrumental variable we draw from Moretti's (2004) approach towards estimating the social returns to education. He proposes that the historical presence of a college be used as an instrument for the relative supply of college graduates. Another proposal is to work with a panel of districts and use a fixed effect estimation to difference out district specific unobservable factors.

Moretti's (2004) idea to use the historical presence of a college as an exogenous predictor of the variation in the stock of highly educated labor across districts can also be applied in the case of the Central European countries (e.g. Jurajda, 2004). Because of limited cross-district labor mobility, the number of college graduates in the district population is to a large extent driven by the presence of a college in this district. Additionally, the majority of public colleges in the Czech Republic and Hungary were established during communism, which makes their presence exogenous to current productivity shocks. Thus, the presence

of bias" (p. 176).

[14] A positive demand shock in the "college" sector makes more graduates work there and thus decreases π^*_{Ckt}. At the same time, it triggers growth in $CollSh_{kt}$. What we observe is a growth in the relative supply of college graduates and a decline in the fraction of them employed in "noncollege" occupations, which creates the impression of a negative relationship between these two.

and/or size of a college[15] in a district as of the end of communism might be a good candidate for an instrument predicting the current stock of college graduates across districts. Although some colleges opened in the 1950's and 1960's were tied to local industries, which casts some doubt on the exogeneity of such instrumental variables, the industrial structure of districts changed during the period of transition and the overall demand for labor has dropped during that time. That is why, while controlling for districts' industrial structure at the end of communism, we can safely use the chosen instruments.[16]

The size and presence of a college in a district as of the end of communism can be used as instruments only in the case of cross-sectional analysis because these instruments do not vary over time. When applying the instrumental variable approach, we are left with a variation in the relative amount of college graduates across districts that is due solely to the historical distribution of colleges and thus is uncorrelated with current district-specific productivity shocks. This should allow for identification of the unbiased cross-district relationship between the relative stock of college graduates and the fraction of them working in "noncollege" occupations.

5 Identifying "college" and "noncollege" occupations

In order to perform the estimations described above, we need to measure the fraction of college graduates employed in "noncollege" occupations. Thus, we need to classify all occupations where college graduates work into "college" and "noncollege" ones. In doing so we follow Gottschalk and Hansen's (2003) approach based on the model presented in Section **??**. This approach exploits the property of the model described by inequality (3), i.e. that wages of college graduates relative to high school graduates are higher in sector 1, the "college" sector. This can be further extended to the situation when there are many different occupations in each sector, but still it holds that in each "college" occupation, the relative productivity

[15] Size of the district's college as of the end of communism is defined as the fraction of the district population holding a college degree in 1991.

[16] *Both presence of a college and size of a college in a district as of the end of communism are strong instruments (correlation with 2001 share of college graduates is 0.63 and 0.85, respectively). Additionally, Sargen's test of overidentifying restrictions suggests that, given the presence of a college in 1991 is exogenous to the model, its size is exogenous as well (p-value = 0.512).*

of college graduates is higher than in each "noncollege" occupation. Consequently, also the relative wages of college graduates are higher in occupations from "college" sector than from "noncollege" sector.

Based on this model, we can distinguish between "college" and "noncollege" occupations once knowing the wage premium paid to college-educated workers over high school-educated workers in each occupation employing both worker types. Gottschalk and Hansen, who perform an occupational classification for the U.S., use a 10% college wage premium as a threshold, i.e., they classify an occupation as "college" when it pays at least 10% premium to highly educated workers.[17] This value, as they justify it, is a bit higher than the lowest estimate of the overall college wage premium in the U.S. as estimated by Katz and Murphy (1992). Taking into account that the overall college wage premium in the Central European countries is significantly higher than in the U.S., we use a higher threshold (15%). Nevertheless, as presented in the next section, the qualitative results are insensitive to the chosen threshold.

Occupations where one type of workers strongly prevails are classified automatically. Gottschalk and Hansen propose to call occupations where more than 90% of workers have higher education as "college" ones. Due to a low fraction of college graduates in the Central European labor market, we lower this threshold to 85%. Additionally, we classify occupations where more than 95% of workers have a high school diploma as "noncollege" occupations.

The procedure of classifying occupations can be described in the following way. For each 3-digit occupation where college graduates constitute between 5% and 85% of all employees, we estimate the following wage equation:

$$\log w_{ik} = \beta_{0k} + \beta_{1k} \cdot exp_i + \beta_{2k} \cdot exp_i^2 + \beta_{3k} \cdot female_i + \phi_k \cdot coll_i + \varepsilon_{ik}, \tag{17}$$

where $\log w_{ik}$ is the logarithm of hourly wage received by an individual worker i in occupation k, exp_i and exp_i^2 are each worker's potential labor market experience (in years) and its square, $female_i$ is a dummy variable indicating a worker's gender and $coll_i$ is a

[17] The same threshold is used by Cardoso (2007) for analysing the Portugese situation and by Grazier et al. (2008) for analysing the British labor market.

dummy variable equal to 1 if worker has a college degree and 0 otherwise.[18] This is a standard Mincerian regression used widely in the literature for identification of returns to different workers characteristics. The parameter used for classification of occupations is ϕ_k, the college wage premium. Occupations for which the hypothesis that $\widehat{\phi_k} > threshold$ (where $threshold$ is initially set at 0.15) can not be rejected are classified as "college" ones. Those for which this hypothesis is rejected are classified as "noncollege". Finally, the occupations where more than 85% of employees are college graduates are classified as "college" occupations and those where less than 5% of employees are college graduates are classified as "noncollege" occupations.

6 Estimation of the influence of college supply on allocation of college graduates across occupations.

6.1 Data description

For the purpose of the empirical analysis we use linked employee-employer data (LEED) from the Czech Republic, Hungary, and Slovakia. The Czech and Slovak national employer surveys, called Information System on Average Earnings (ISAE), contain data on a representative sample of over 3500 (1500) firms with more than 10 employees. This results in about 1.3 (0.5) million worker level observations a year in the Czech (Slovak) Republic. The Hungarian Wage and Employment Survey (WES) investigates all tax-paying legal entities with double-sided balance sheets that employ at least 10 employees. Nevertheless, individual workers are selected into the sample based on their date of birth: production workers are included if their birth date falls on either the 5th or the 15th of any month, and non-production workers if it falls on the 5th, 10th, or 15th of a month. This results in about 150 thousand worker level observations a year for Hungary.

Both the ISAE's and WES databases include detailed demographic information for workers, as well as variables describing the size and location of the firm where they are employed.

[18] The sample used for classification of occupations contains all college and high school educated workers not older than 35. The sample choice is discussed in more detail in the next section.

Worker variables include gender, age, highest education level, and occupation. We run our analysis using samples of workers belonging in two education groups: college and high school graduates. Occupations in which they are employed are defined at the 3 digit level.[19] For the purposes of our estimation, occupations with a very small number of observations were merged based on their definitions to give categories with at least 70 workers. The wage measure used in the Czech and Slovak samples is the hourly wage which is defined as the average pay per hour during the first quarter of a year; while in the Hungarian sample we use the net monthly wage of each worker, calculated from the monthly base wage and the previous year's additional benefits and bonuses. As the Hungarian dataset contains only full-time workers, the wage measures should be comparable.

The data on the 2001 and 199 (1990) educational composition of the population for each NUTS4[20] district come from national Censuses and were provided by the respective statistical offices. In the Czech and Slovak Republics there are about 70 NUTS4 districts. Hungary is divided into 173 NUTS4 districts that were merged based on geographical locations into 60 larger districts in order to ensure adequate sample size in each district.

6.2 Cross-sectional Estimation at the District Level

This section presents the estimates of the relationship between the relative number of college graduates in the population and the probability of college graduates to work in "noncollege" occupations, as described by equation (16). As shown in Table 1, this analysis supplies some evidence that the productivity spillover from a high concentration of skills is strong enough to create improved employment possibilities for college graduates in districts where their stock is relatively high. The table reports the estimates of θ_1 obtained using different models (OLS and IV), and two alternative thresholds for defining "college" occupations (10% and 15%).

[19] Both in the Czech and Slovak datasets occupations are coded according to a local system (KAMZ) which follows the International Standard Classification of Occupations (ISCO). In the Hungarian dataset occupations are coded according to a FEOR system, which corresponds to ISCO only at the 1-digit level. Nevertheless, this system follows similar rules.

[20] NUTS is the European Union's Nomenclature of Units for Territorial Statistics. Level 4 is the most detailed division into regions with less than 800 000 citizens each.

The estimations are repeated with and without the capital cities districts because they are characterized by incomparably large share of college graduates in the local populations and high concentration of businesses.

Table 1: Determinants of the share of college graduates in "noncollege" occupations - using 15 percent threshold - across Czech districts in 2001

	OLS			IV		
	Czech Republic	Hungary	Slovakia	Czech Republic	Hungary	Slovakia
10% threshold						
with	-1.241**	-0.495**	-0.128*	-0.890	-0.628*	0.017
capital city	(0.030)	(0.030)	(0.088)	(0.150)	(0.052)	(0.437)
excluding	-1.250**	-0.528**	-0.149*	-0.897	-0.688**	0.016
capital city	(0.028)	(0.025)	(0.078)	(0.146)	(0.047)	(0.382)
15% threshold						
with	-0.985*	-0.471*	-0.281	-0.737	-0.643*	0.008
capital city	(0.065)	(0.051)	(0.101)	(0.203)	(0.055)	(0.379)
excluding	-0.994*	-0.498**	-0.285*	-0.746	-0.691*	-0.069
capital city	(0.065)	(0.049)	(0.085)	(0.201)	(0.055)	(0.325)

Notes: The dependent variable is individual young college graduate's probability of working in a noncollege occupation. *CollShare* is the 2001 share of college graduates in a respective district's young population; as an IV for this variable, we use the share of college graduates in district population as of the end of communism (1991 or 1990). Young workers are defined as being not older than 40. P-values are in parentheses.

Table 1 indicates that the estimates of the influence of the relative number of college graduates in a district population on the fraction of them working in "noncollege" occupations are significant and negative in all three countries when the OLS estimation method is applied. These results are, however, biased downwards due to the simultaneity in the determination of these two variables. Thus, we should expect the true relationship to be different. Indeed, when instrumenting the 2001 share of college graduates in the district population with the same measure as of the end of communism, estimates closer to zero are obtained for the Czech and Slovak Republics, while more negative (and significant) estimates are obtained for Hungary. In the Czech Republic the relationship between the relative stock of college graduates in the district population and the fraction of them working in "noncollege" occupations is estimated to be different from zero with only 85% confidence and is not statistically different from zero at any reasonable confidence level in Slovakia. Nevertheless, it is not estimated to be positive, which would be the expected result when no spillover effects are present.[21] Actually, the economic significance of the coefficient by $CollShare$ is quite strong – a one percentage point increase in the share of college graduates in the local labor market is estimated to cause a 0.9 (0.6) percentage point decrease in the fraction of college graduates working in "noncollege" occupations in the Czech Republic (Hungary). This gives us some evidence to support the hypothesis that a larger number of college graduates attracts advanced technologies and in this way improves the situation of highly educated workers in the district labor market.

Let us note that instrumenting the 2001 share of college graduates in the district population with the same measure as of the end of communism in Hungary had opposite effects on the estimates of θ_1 than in the Czech and Slovak Republics. This could be the signal that the chosen instrument is not excluded from equation 16 for Hungary, for example because the relative productivity of college graduates became higher in regions which were initially rich in skilled labor. Alternatively, this result could be obtained in the situation when the OLS estimates are biased upwards for Hungary, which would be observed if the government decides to expand higher education in disadvantaged districts. Distinguishing between these two possibilities will be the next step of our research.

7 Conclusion

In this study we argue that the fraction of college graduates employed in "noncollege" occupations offers a useful measure for investigating forces shaping the labor market. Analysis of

[21]Recall that, according to equation (15), $\delta_2 > 0$. Thus a non-positive estimate of $\theta_1 = \delta_1 + \delta_2$ implies that $\delta_1 < 0$, i.e. that the spillover effect exists.

the evolution of this measure over time in the U.S. (Gottschalk and Hansen 2003), Portugal (Cardoso 2007), UK (Grazier 2008)and the Czech Republic (Gebicka 2010) reveals a consistent pattern. In every country the fraction of college graduates employed in "noncollege" occupations has been decreasing over time despite a significant growth in the relative number of college educated workers in the labor market. This phenomenon could be driven by two forces: (1) exogenous technological shocks simultaneously triggering shifts in the demand for and supply of college graduates, or (2) a higher number of college graduates attracting advanced technologies and thus endogenously shifting the demand for skilled workers.

These forces are not mutually exclusive; most probably they act simultaneously. Nevertheless, from the policy point of view it is important to know how strong the endogenous effect is as compared to the exogenous effect. In the absence of the endogenous effect, college enrolments should reflect the trend in technological progress of the economy; while the existence of this effect implies that increasing the educational attainment of the local population could be used as a tool to attract advanced technologies and increase the skill bias of the economy.

Results presented in this paper confirm the presence of a negative influence of the number of skilled workers on the fraction of them working in "noncollege" occupations across NUTS-4 districts of the Central European countries. This is in line with the findings of Acemoglu (2003), who shows that a high supply of skilled labor shifts the skill bias of the local economy. Thus, the findings of this paper suggest that in the long run, districts should be able to positively stimulate their labor markets by providing higher education to a larger fraction of their population (explanation 2).

References

[1] Acemoglu D. (2002) Technical change, inequality, and the labor market, *Journal of Economic Literature*, 40(1): 7-72.

[2] Acemoglu D. (2003) Patterns of skill premia, *The Review of Economic Studies*, 70(243):199-230.

[3] Acemoglu D., Angrist J. (2000) How large are human capital externalities? Evidence from compulsory schooling laws, NBER Macroeconomics Annual, MIT Press: Boston.

[4] Bound J., Green J., Kezdi G., Turner S. (2004) Trade in university training: cross-state variation in the production and stock of college-educated labor, *Journal of Econometrics*, 121(1-2): 143-173.

[5] Bound J., Johnson G. (1992) Changes in the structure of wages in 1980's: An evaluation of alternative explanations, *American Economic Review*, 82: 371-92.

[6] Brunello G., Comi S., Lucifora C. (2000) The college wage gap in 10 European countries: Evidence from two cohorts, *IZA Discussion Paper* No. 228.

[7] Card D., Lemieux T. (2001) Can falling supply explain the rising return to college for younger men? A cohort-based analysis, *The Quarterly Journal of Economics*, 116(2): 705-746.

[8] Cardoso A.R. (2007) Jobs for young university graduates, *Economic Letters*, 94(2): 271-277.

[9] Cseres-Gergely Z. (2002) Migration, migrants and the role of economic incentives to migrate in Hungary 1990-1999. Budapest, MTA-KTK.

[10] Eurostat (2009) http://nui.epp.eurostat.ec.europa.eu/nui/show.do?dataset=lfsa_pgaed&lang=en

[11] Fidrmuc, J. (2004) Migration and regional adjustment to asymmetric shocks in transition economies, *Journal of Comparative Economics,* 32: 230–247.

[12] Filer R., Jurajda S., Planovsky J. (1999) Education and wages in the Czech and Slovak Republics during transition, *Labour Economics*, 6(4): 581-593.

[13] Flabbi L., Paternostro S., Tiongson E.R.(2008) Returns to education in the economic transition: A systematic assessment using comparable data, *Economics of Education Review*, 27(6): 724-740.

[14] Fortin N.M. (2006) Higher education policies and the college wage premium: Cross-state evidence from the 1990's, *American Economic Review*, 96(4): 959-987.

[15] Galasi P. (2005) Reallocation of Workers with the Higher Education Diploma 1994-2002, in K. Fazekas and G. Kezdi (eds.) "The Hungarian Labour Market Review and Analysis 2005", Budapest, Institute of Economics and Hungarian Employment Foundation: pp. 104-122.

[16] Galasi P. (2007) Earnings of higher-education graduates: the role of education, type of education and under/over-education, in K. Fazekas and G. Kezdi (eds.) "The Hungarian Labour Market. Review and Analysis 2007", Budapest, Institute of Economics and Hungarian Employment Foundation: pp. 104-122

[17] Gebicka B. (2010) College Degree Supply and Occupational Allocation of Graduates – the Case of the Czech Republic, *CERGE-EI Working Paper No. 407*, 20(1): 41-87.

[18] Grazier S., O'Leary N., Sloane P. (2008) Graduate employment in the UK: An application of the Gottschalk-Hansen model, *IZA Discussion Paper* No. 3618.

[19] Gottschalk P., Hansen M. (2003) Is the proportion of college workers in noncollege jobs increasing? *Journal of Labor Economics*, 21(2): 449-471.

[20] Lannert J. (2005) Facts on Expansion of Education, The Hungarian Labour Market Review and Analysis 2005, in K. Fazekas and G. Kezdi (eds.) "The Hungarian Labour Market Review and Analysis 2005", Budapest, Institute of Economics and Hungarian Employment Foundation: pp. 104-122.

[21] Jurajda S. (2005) Czech returns to schooling: does the short supply of college education bite? *Czech Journal of Economics and Finance (Finance a úvěr)*, 55(1-2): 83-95.

[22] Jurajda S. (2004) Are there increasing returns to local concentration of skills? Evidence on wages and returns to education in transition, *mimeo*

[23] Jurajda S., Terrell K. (2009) Regional unemployment and human capital in transition countries, *Economics of Transition,* 17: 241–274,

[24] Katz L., Autor D. (1999) Changes in the wage structure and earnings inequality, in: Ashenfelter O., Card D. (eds.), Handbook of Labor Economics, Amsterdam.

[25] Katz L., Murphy K. (1992) Changes in relative wages, 1963-1987: Supply and demand factors, *The Quarterly Journal of Economics*, 107(1): 35-78.

[26] Kertesi, G., Kollo, J. (2005) The Expansion of Higher Education, High-Skilled Unemployment, and the Returns to College-University Diploma, *Budapest Working Papers on the Labour Market* 2005/3

[27] McGuinness S. (2006) Overeducation in the labor market, *Journal of Economic Surveys*, 20(3): 387-418.

[28] Moretti E. (2004) Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data, *Journal of Econometrics*, 121(1-2): 175-212.

[29] Munich D., Svejnar J., Terrell K. (2005) Returns to human capital under the communist wage grid and during the transition to a market economy, *Review of Economics and Statistics,* 87(1): 100-123.

[30] Pryor F., Schaffer, D. (1997) Wages and the university educated: a paradox resolved, *Monthly Labor Review*, 120(7): 3-14.

[31] Wheeler C.H. (2001) Search, sorting, and urban agglomeration, *Journal of Labor Economics*,19(4): 879-899.