# College degree supply and occupational allocation of graduates - the case of the Czech Republic

Barbara Gebicka*

CERGE-EI

August 2009

**Abstract**

There is much research on skill-biased changes in labor demand and the simultaneous growth in the number of college graduates. A key question is whether the two proceed in a balanced fashion. In this paper, I apply the technique developed by Gottschalk and Hansen (2003) to identify the skill requirements of occupations and to study the share of college graduates in noncollege occupations in a country experiencing a dramatic expansion of higher education: the Czech Republic. Comparing districts with different education structures of population suggests that the higher is the stock of graduates, the better is the efficiency of matching them with college occupations. Nevertheless, no such effect is found for within-districts changes in the amount of graduates, possibly due to the short time span of data used. These findings are consistent with skill-complementing capital locating in places relatively abundant in skills, which needs time to be realized. This suggests that the supply of college seats should not only be a response to the observed level of demand for skills but also a tool for attracting technologically advanced industries and improving the employment situation of skilled labor in the long term.

*I am grateful to Štěpán Jurajda for his guidance and advice. I also thank Alena Bičáková, Daniel Münich, Randall Filer and CERGE-EI Brown Bag Seminar participants for helpful hints and comments. All remaining errors are mine. This research has been supported by the World Bank Research Fellowship grant. E-mail: barbara.gebicka@cerge-ei.cz

# 1 Introduction

Central-European economies inherited tiny and rigid higher education systems from the communist times. They managed to provide college education to just about 10% of adult population, while the developed economies had on average 25% of college graduates in the early 1990's. This under-supply of skills was reflected in high returns to college education during the first decade of transition (reference). The following rapid expansion of higher education significantly increased the educational attainment in Central European countries. In the Czech Republic, for example, college enrollment rates more than doubled during the last 15 years and the fraction of college graduates in adult population increased to 15%. At the same time, the post-communist economies are exposed to capital inflow, mainly in the form of FDI, and import of advanced technologies from the developed countries. This shifts the demand for highly educated labor even further. Although the fraction of college graduates in these economies is sharply growing, the returns to higher education do not fall (source).

Economic literature recognizes several channels influencing the demand for skilled workers. First, it might be shifting exogenously[1] - due to a natural catch-up after opening-up of the post-communist economies coupled with the skill biased technological change (SBTC). Second, shifts in the demand for skilled labor might be accelerated by endogenous effects - when the presence of many college graduates in an economy attracts technologically advanced firms and increases innovativeness. The question addressed in this paper is whether the latter effect influences significantly the labor market in a post-communist economy, the Czech Republic. I test whether a higher number of college graduates across Czech districts attracts firms using advanced technologies. If present, this effect would be reflected in more workplaces for college graduates and thus better utilization of their skills than would be expected in the presence of exogenous demands shifts only.

Identification of a positive relationship between the number of college graduates and creation of skill-intensive workplaces would inform policy decisions concerning provision of

[1] This should be understood as shifts independent of the labor force structure.

higher education. In the absence of this endogenous effect, college enrolments should just reflect the trend in technological progress of the economy; while existence of this effect means that increasing the educational attainment of local population could be used as a tool additionally attracting advanced technologies and increasing the skill bias of the economy. This is of special relevance in Central-European countries which still lag behind the Western economies in terms of technological development. Additionally, the policy relevance of this research lies in the fact that higher education systems in the post-communist countries are largely state-funded and thus provision of higher education is a public policy decision. Knowing the channels affecting the demand for college-educated labor would facilitate decision-making concerning the extent of higher education expansion.

The primary concern of this research is to recognize whether the inflow of technologically advanced firms, which provide employment for college graduates, is affected by the concentration of skilled labor in a local economy. I investigate the relationship between the number of college graduates in Czech districts' population and the kind of occupations they are employed in to find out that in the long run high concentration of skilled labor has a positive effect on the level of skill utilization in this district. This gives some evidence to believe that the presence of many college graduates attracts advanced technologies. Thus, providing higher education to higher fraction of local population could be used as a policy enhancing technological development of the economy.

As a measure of the level of college-gained skills utilization, I use the propensity of a college graduate to work in a "college" occupation. Classification of occupations into "college" and "noncollege" has been proposed by Gottschalk and Hansen (2003). They develop a model where "college" occupations are characterized by a higher relative productivity of college to high school graduates than "noncollege" occupations. Then they use the college-high school wage premium paid by individual occupations as a proxy for the relative productivity and classify occupations as "college" if the wage premium they pay exceed certain threshold. This approach and its limitations are discussed in Section 3 of the paper.

The econometric analysis of the relationship between the fraction of skilled workers in an economy and the propensity of a college graduate to work in a "college" occupation is done in both the cross-district and panel of districts dimension. In the cross-district context, I identify the causal relationship using the historical educational structure of local populations as a skill supply shifter exogenous to current labor demand conditions. This allows to identify the long-run effect. The short-run relationship is identified by using districts fixed effects and by including a proxy for year-to-year demand shifts - the Katz and Murphy demand shift index. Interestingly, the long-run and short-run effets go in opposite directions. While in the cross-section I identify a positive influence of high concentration of skills on the propensity of a college graduate to work in a "college" occupation, in the panel of districts analysis this relationship is found to be negative. This suggests that the presence of many college graduates in an economy can attracts technologically advanced firms, but this effect needs time to be realized.

This paper falls in the context of broad U.S. based research focusing on the determinants of the (relative) demand for college-educated labor. For example, Moretti (2004) finds 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 which implies existence of positive productivity spillovers from spacial concentration of skills and suggests that a large number of college graduates in a labor market can trigger a shift in demand for them. Similar conclusions are reached by Acemoglu (2002, 2003) in his theoretical analyses of the demand and supply of skills. Somewhat different results are, however, presented by Fortin (2006). She shows that an increasing production of college graduates in a given state lowers the college-high school wage gap. This could be interpreted in the way that an increase in the number of college graduates does not trigger enough shift in the demand for them to compensate for the movement along the downward sloping relative demand curve. Nevertheless, this finding does not contradict earlier research. A shifting supply of college graduates can result in a decline of their relative wages, but at the same time their absolute wages might grow. In the context of Central-European economies it is worth to mention a study by Jurajda (2004), who tests for existence of productivity spillovers from high concentration of college graduates across Czech districts. Interestingly, he finds no evidence in support of such spillovers. Czech college graduates' wages appear to be insensitive to their concentration.

This finding does not, however, exclude the existence of potential positive effects of expanding higher education. The Czech labor market could react by offering more workplaces for college graduates,[2] what is investigated in more details in this paper.

The rest of this article is organized in the following way. The next section describes higher education in the Czech Republic. Then, the theoretical and empirical models of college and high school graduates' allocation across different occupations are described. The data description follows. Finally, I present and discuss the estimation of the causal relationship between the relative stock of college graduates and the fraction of them working in "noncollege" occupations. The last section concludes.

## 2 The Czech Republic

The analysis presented in this paper focuses on employment of college graduates in the Czech Republic. This country is particularly interesting because of several aspects concerning the organization of tertiary education. The majority of Czech public universities have been established during the communist times and their restructuring started only in the 1990's. This is visible not only in the significant increase in enrollment rates in the last decade but also in the structure of study fields offered (CSO, 2008). The restructuring of higher education has not, however, dealt with sources of universities financing. Tertiary education in the Czech Republic is still largely state-funded.

The growth in college enrollment and the resulting increased inflow of graduates is changing the educational structure of the Czech population. The fraction of college graduates in prime-age population has grown from 11% in 2000 to 14% in 2008 with an average annual growth rate of over 3% (Eurostat 2009). Despite these changes, the fraction of the prime-age population with higher education is still very low in the Czech Republic as compared to other countries. The OECD average fraction of college graduates among prime-age population was 27% in 2006 (OECD 2009) with the U.S. having the highest number (39%).

[2] For example, Jurajda and Terrell (2009) find that FDI flows to regions characterized by higher concentration of college educated.

International comparison suggests that the Czech Republic will be undergoing further expansion of higher education in the years to come in order to catch up with other countries. Thus, it is important to know how these changes shape the labor market.

Another interesting characteristic of the Czech Republic is the cross-district[3] diversity in the educational structure of its population. There are significant differences in the fraction of the adult population with tertiary education and the rates of growth in this measure as presented in Figure 1. Additionally, readers should notice that districts which had a college established by the end of communism are characterized by significantly larger shares of highly-educated population, but do not differ from other districts in terms of growth rates. This property is used as an exclusion restriction when identifying the influence of the relative supply of college graduates on their fraction working in "noncollege" occupations, what is discussed in detail in Section 4.

[Figure 1 here]

Finally, tertiary education in the Czech Republic is largely state funded. The supply of places in tuition-free colleges (which is significantly lower than the demand for them) is determined by the funds allocated by policy makers. In other words, enrollment in higher education institutions constitutes a public policy decision. This is almost directly translated into the future number of college graduates in local labor markets because of low cross-district migration in the Czech Republic.[4] Although low migration within the Czech Republic has been already documented by Fidrmuc (2004), let me present here an alternative evidence on this phenomenon. To illustrate what is the extent of cross-district migration among college graduates in the Czech Republic, I compare the district-specific numbers of college graduates in two 5-year age cohorts (30-34 and 35-40) as recorded by the 1991 Census,

[3] Districts are NUTS-4 (Nomenclature of Statistical Territorial Units of the European Union) regions with population of less than 150,000 individuals.

[4] Bound et al (2004) show that the relationship between the production and stock of college graduates in US states is weak and thus state-specific educational policies might not have the desired effect on the labor market. This, however, appears not to be the case in the Czech Republic.

with corresponding cohorts in the 2001 Census.[5] This comparison is presented in Table 2 in the Appendix. The average change in the number of college graduates in these two cohorts between 1991 and 2001 was 10%. One can see in the presented table that within district changes only slightly differ from this number. This might suggest that cross-district migration of college graduates is very low. There are two outlying districts experiencing a decrease in the number of college graduates (Jindrichuv Hradec and Sumperk) and one district experiencing a very large increase in this number (Uherske Hradiste). These districts will be dropped from the final analysis to avoid misspecification.

Focusing on a country with significant district-level differences in the educational structure of the population driven by public policy decisions enables one to investigate how these decisions influence the situation of graduates in the labor market. It is especially interesting to see if, in the districts with higher skill endowment and/or where higher education is expanding more rapidly, it is easier or more difficult for college graduates to find employment that takes advantage of their skills. This analysis is of particular policy interest because it reveals whether in this setting expansion of higher education can improve employment possibilities of college graduates (and thus their skill usage) by attracting firms willing to employ them.

# 3 Theoretical framework

In this paper I estimate the influence of variations in the relative stock of college graduates on their occupational allocation. For simplicity, two broad groups of occupations are considered: "college" and "noncollege" occupations. In line with the existing literature on the topic, I identify two main forces that determine this relationship. First is the movement along the downward sloping demand curve which causes a decrease of college graduates' wages when the number willing to supply labor to a given type of occupations (say, the technologically advanced ones) increases. This, in turn, makes other occupations relatively more attractive, which finally results in reallocation of college graduates across occupations. The second force

[5] The districts of Prague and Brno, the outliers in the number of college graduates, have been removed from this analysis.

is a productivity spillover from having many highly educated workers. Such productivity spillovers have been already identified in the literature (Moretti 2004, Acemoglu and Angrist 2000), but never in the context of allocation of workers across occupations. The spillover idea is the following. The presence of many college graduates in the labor market makes it possible for them to cooperate and thus take advantage of each other's skills, which increases their overall productivity. Additionally, the presence of many college graduates attracts technologically advanced firms because it is easier for them to find the employees they need among the pool of highly educated. Consequently, there are more firms offering employment in "college" occupations and wages paid by them are relatively higher, which attracts more college graduates to work there - an opposite reallocation than the demand effect described before.

In this section I model the channels through which the educational structure of the labor market influences college graduates' allocation across "college" and "noncollege" occupations. First, I present a simple demand-supply framework explaining how the fraction of college graduates working in "noncollege" occupations depends on their relative stock in a closed, competitive economy. Further on, I extend my theoretical analysis to allow for a set of economies across which labor and capital are free to move. This leads to the formulation of an econometric model which I estimate in the next section.

## 3.1 Single closed economy

This section outlines a modification of the Gottschalk and Hansen's (2003) model allowing for identification of the direct relationship between the relative stock of college graduates and the fraction of them working in "noncollege" occupations. This model 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. They have the following production functions:

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

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

where $L_{Cj}$ and $L_{Nj}$ are the amounts of college- and high school-educated labor in sector $j$, and $\alpha_{ij}$ are productivities of labor type $i$ in sector $j$. 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. wages of college graduates relative to high school graduates are higher in sector 1, the "college" sector. This property will be further used to distinguish between "college" and "noncollege" occupations.

To complete the model, the supply of different labor types to both sectors needs to be specified. Following Gottschalk and Hansen (2003), I 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) In the original model, the relationship between the sector-specific supply functions and the total number of college and high school graduates in the labor market is not explicitly shown.[6] 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 my 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:

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

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

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

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

$$L^S_{N2} = L_N - L^S_{N1}, \tag{7}$$

where $L_C$ and $L_N$ are the total amounts of college and high school graduates in the labor market, and $\beta_i$ and $\lambda_i$ are the aggregate preference parameters of workers of type $i$.

Together, equations (3)[7] 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 is that in equilibrium there are some college-educated workers employed in both sectors. In the further analysis I am interested in the fraction of college graduates working in the "noncollege" sector, which is defined as

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

In equilibrium this is equal to

$$\pi^*_C \equiv 1 - \frac{L^*_{C1}}{L_C} = f\left(L_C, L_N, \alpha_{C1}, \alpha_{N1}, \alpha_{C2}, \alpha_{N2}\right), \tag{9}$$

and depends on the supply conditions, i.e., the total amount of each labor type in the economy ($L_i$) and demand conditions, i.e., labor productivities ($\alpha_{ij}$). The sign of the relationship of main interest, i.e. how the equilibrium allocation of college graduates across occupations depends on the structure of the labor market ($L_C/L_N$) and on the extent of the skill bias of technology ($\frac{\alpha_{C1}}{\alpha_{N1}}$), is derived below.

First, I analyze how the equilibrium allocation changes when SBTC happens in the "college" sector, i.e., when $\frac{\alpha_{C1}}{\alpha_{N1}}$ grows and all other parameters 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 is 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

[7] Equation (3) actually consists of 4 equations: $w_{C1} = \alpha_{C1}\frac{\partial F_1(\cdot)}{\partial L_{C1}}$, $w_{N1} = \alpha_{N1}\frac{\partial F_1(\cdot)}{\partial L_{N1}}$, $w_{C2} = \alpha_{C2}\frac{\partial F_2(\cdot)}{\partial L_{C2}}$, and $w_{N2} = \alpha_{N2}\frac{\partial F_2(\cdot)}{\partial L_{N2}}$.

higher wages for college graduates in both sectors, but wages in sector 1 increase more as compared to the initial level. This makes new $\frac{w_{C1}^*}{w_{C2}^*}$ higher than the initial one and thus new $\pi_C^*$ lower than the initial one. To sum up,

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

Next, let me analyze what happens when the relative stock of college graduates in the labor market ($L_C/L_N$) 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 the result wages of all labor types in the "college" sector fall. In sector 2 wages fall as well, but less dramatically, as long as $\frac{\alpha_{C2}}{\alpha_{N2}} > 1$. 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 finally 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 new $\frac{w_{C1}^*}{w_{C2}^*}$ lower than the initial one and thus new $\pi_C^*$ higher than the initial one. To sum up,

$$\frac{\partial \pi_C^*}{\partial \left(L_C/L_N\right)} > 0. \tag{11}$$

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:

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

Assuming that the relationship is approximately log-linear[8] and other parameters do not vary, I can write it in the following form:

$$\ln\left(\pi_C^*\right) = \gamma_0 + \gamma_1 \ln\left(\frac{L_C}{L_N}\right) + \gamma_2 \ln\left(\frac{\alpha_{C1}}{\alpha_{N1}}\right), \tag{13}$$

[8] The model outlined in this Section has no closed form solution. Therefore, I have to approximate its functional form.

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

According to the model presented above, the relationship between $\frac{L_C}{L_N}$ and $\pi_C^*$ is positive. However, this model does not take into account the spillover effects from high concentration of skills described at the beginning of this section. Let me now introduce such spillovers to the model to show that they can alternate the relationship of interest. A general representation of productivity spillovers commonly used in the literature is in the form of productivity being an increasing function of total skills (e.g., Acemoglu and Angrist 2002, Moretti 2004). In this paper I use a simple linear relationship:

$$\ln\left(\frac{\alpha_{C1}}{\alpha_{N1}}\right) = \alpha + \delta \ln\left(\frac{L_C}{L_N}\right), \tag{14}$$

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

$$\ln\left(\pi_C^*\right) = \gamma_0 + (\gamma_1 + \gamma_2\delta)\ln\left(\frac{L_C}{L_N}\right) + \gamma_2 \cdot \alpha. \tag{15a}$$

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 ($\gamma_1$) 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 above analysis holds for a single closed economy where college graduates are produced and employed locally. It is presented here to explicitly show the forces behind workers' allocation across "college" and "noncollege" occupations. In this paper, however, I analyze a set of districts across which workers and firms can freely move. This calls for expanding the model.

## 3.2 A set of free-trading economies

The model describing the allocation of workers across occupations and districts is based on Roback (1982, 1988) and extended to include two firm types. For simplicity, let me assume that there are two districts ($A$ and $B$) within one country and both workers and firms are free to move between them. As before, there are two firm types– "college" and "noncollege" ones. There are also two worker types – holders of college and high school diploma. Firms produce a uniform nationally traded good by hiring both worker types and using district-specific land. Workers earn wages and spend them on the uniform good and land for living purposes. Thus, firms prefer low wages and low rental price of land while workers prefer high wages and low rental price of land. Because workers are free to work in whichever region, in equilibrium they must achieve the same utility level no matter where they choose to locate. The same logic applies to firms – in equilibrium they bear the same unit costs in each district.

Within this setup the number of college graduates in one of the districts (say, district B) would be higher if it offers some exogenously determined amenities which highly-educated workers are willing to trade for lower wages and higher rents. These can be cultural amenities attractive for college graduates or the presence of a college with the assumption that graduates prefer living in the district where they have graduated.[9] Under these conditions, rents in city B are higher than in city A, college-educated workers' wages are lower, and a higher fraction of college graduates work in the "noncollege" sector in city B than in city A.

Let me now introduce spillovers to this model. If these effects exist, a shift in the supply of college graduates in city B increases their relative productivity in the "college" sector[10] in that city and thus allows "college" firms to pay higher wages and higher rents to achieve the same unit costs. This effect magnifies the difference in rents between cities B and A. At the same time this effect reduces the drop in wages in the "college" sector in city B, making all or at least some college graduates stay there and possibly attracting additional ones. Thus, the spillover effect reduces or might even reverse the direct effect and make the final ratio of

[9] The latter is partially confirmed by Figure 3, where we observe that districts with a well established college have higher fraction of college graduates in their population, and by low mobility of college graduates, as documated in Table 2.

[10] The "noncollege" sector is not affected by productivity spillovers in a biased way. The increased supply of college graduates in district B can only increase productivities of college- and high school-educated workers in the "noncollege" sector proportionally.

college wages in the "college" and "noncollege" sector in city B high enough to decrease the fraction of college graduates working in the "noncollege" sector.

The mechanisms described above can be summarized by he following district-level equation:

$$\ln(\pi_C^*) = \gamma_0 + (\delta_2 + \delta_1)\ln\left(\frac{L_C}{L_N}\right) + \varepsilon, \tag{16}$$

where $\delta_1$ is the spillover effect and $\delta_2$ is the demand effect. $\varepsilon$ captures all other factors that influence the allocation of college graduates across occupations. The expected signs of the above parameters are $\delta_1 < 0$, and $\delta_2 > 0$. The direction of their joint influence, i.e., the sign of $\theta_1 \equiv \delta_2 + \delta_1$, can not be theoretically predicted. The goal of this paper is to estimate the parameter $\theta_1$ to find out whether positive or negative effects prevail in the influence of the relative stock of college graduates on their allocation across occupations.

# 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 me note that equation (16) accommodates an implicit assumption that aggregate preference parameters of workers 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 college education (who, on average, are 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, I formulate an econometric model on the individual rather than aggregate level, i.e., I 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 (17). This model can be thought of as a disaggregated version of equation (16).

$$\text{Prob}(nocollege_{ikt}) = \gamma_0 + \mathbf{X}'_{ikt}\boldsymbol{\theta}_0 + \theta_1 \ln\left(\frac{L_C}{L_N}\right)_{kt} + \mathbf{Y}'_{kt}\boldsymbol{\theta}_2 + \varepsilon_{ikt}, \tag{17}$$

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, $\ln\left(\frac{L_C}{L_N}\right)_{kt}$ 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 $\varepsilon_{ikt}$ represents the individual, time and district specific unobservable determinants of college graduates' allocation across occupations. The parameter of main interest is $\theta_1$; it describes the causal relationship between the relative amount 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. I 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 highly educated.[12]

The source of identification used to estimate $\theta_1$ is the variation in the fraction of highly educated adults across and within Czech districts' population and the simultaneous variation

[11] Ideally, the above should be modeled as a choice between three alternatives: working in a "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] I 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.

in the proportion of college graduates working in "noncollege" occupations in these districts. Because of the two-level structure of the variables,[13] the precision of $\widehat{\theta}_1$ might be significantly downward-biased if estimating the model (17) by standard methods. Simple clustering would not improve the situation because of a limited number of clusters (districts). As Donald and Lang (2007) show, standard errors of variables that are constant within a group (here within a district in a given year) "*are asymptotically normally distributed only as the number of groups goes to infinity.*" (p. 221) The same authors propose a two-step procedure to get over this problem. I follow this procedure by first estimating the propensity of individual college graduates to work in noncollege occupations as a function of their individual characteristics and district-time dummies. In the second step I perform a weighted least squares (WLS) regression of the estimated parameters by district-time dummies on district-time characteristics, where the variance of the estimated parameters by district-time dummies is used as the weighting factor This approach can be summarized in the following way:

$$1^{\text{st}} \text{ step:} \qquad \text{Prob}(nocollege_{ikt}) = \delta_0 + \mathbf{X}'_{ikt}\boldsymbol{\delta}_1 + \mathbf{TD}'_{kt}\mathbf{d} + \boldsymbol{\xi}_{ikt}, \tag{18a}$$

$$2^{\text{nd}} \text{ step:} \qquad \widehat{d}_{kt} = \gamma_0 + \theta_1 \cdot \ln\left(\frac{L_C}{L_N}\right)_{kt} + \mathbf{Y}'_{kt}\boldsymbol{\theta}_2 + \varepsilon_{kt}, \tag{18b}$$

where $\mathbf{TD}'_{kt}$ is a vector of year-district dummies, $\boldsymbol{\xi}_{ikt}$ captures unobservable individual characteristics and $\varepsilon_{kt}$ represents the time and/or district specific unobservable determinants of college graduates' allocation across occupations.

Unfortunately, an omitted variable problem appears when estimating equation (18b) by WLS.[14] Some of the factors captured by the error term might bias the estimate of $\widehat{\theta}_1$ because of being correlated 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

[13]The dependant varable is on individual level, while the explanatory variable of main interest is on group (district) level.

[14]There might appear also an omitted variable bias when estimating equation (18a) if workers sort into cities according to their unobservable abilities. In this case, $\mathbf{TD}'_{kt}$ and $\boldsymbol{\epsilon}_{ikt}$ are correlated, which influences the estimated of $d_{kt}$. This could be addressed by controlling for workers' fixed effects. The data used in this study do not have a repeated cross-section structure, which does not allow for using this approach. Nevertheless, this should not be a big problem - Moretti (2004) shows that omitted "*individual characteristics are not a major source of bias*" (p. 176).

labor. Both time and district specific productivity shocks might partially drive the variation in the stock of college graduates. For example, an expansion of hi-tech industry in one district attracts highly educated workers to move there or observation of country-wide SBTC motivates more people to get higher education. This is why I expect $cov(\varepsilon_{kt}, \ln \left(\frac{L_C}{L_N}\right)_{kt}) \neq 0$. The intuitive sign of this correlation is positive (i.e., positive productivity shocks induce a higher fraction of college graduates), thus the WLS estimates of the relationship from equation (18b) would be biased downwards.[15]

Endogeneity of the fraction of population with a college degree can be addressed 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 I draw from Moretti's (2004) approach towards estimating the social returns to education. He proposes to use historical presence of a college 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 idea of using historical presence of a college as an exogenous predictor of the variation in the stock of highly educated labor across districts can be applied also in the case of the Czech Republic (e.g. Jurajda, 2004). Because of limited cross-district labor mobility, as discussed in Section 2, 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 have been established during communism, which makes their presence exogenous to current productivity shocks. Thus, the presence and/or size of a college[16] 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

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

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

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 has 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, I can safely use the chosen instruments.

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, I am left with the variation in the relative amount of college graduates across districts that is only due to the historical distribution of colleges and thus is uncorrelated with current district-specific productivity shocks This should thus 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.

Working with a panel of districts allows for identification of the influence of changes in the relative supply of college graduates on their allocation between "college" and "noncollege" occupations. It also allows me to use a fixed-effect estimation approach and difference-out the time-constant district-specific demand shifters. In this way I eliminate the endogenous effect coming from the correlation of district-specific time-constant unobservables and the relative stock of college graduates in a district's population. Nevertheless, there still can be time-varying factors influencing the changes in the relative number of college graduates. Inclusion of a proxy for a time-district specific demand factors - the Katz and Murphy demand shift index[17] - would remove some of the unobservable demand from the error term and minimize the bias of $\widehat{\theta}_1$.

[17] More information about the Katz and murphy demand shift index can be found in Katz and Murphy (1992) and Moretti (2004).

# 5 Identifying “college” and “noncollege” occupations

In order to perform the estimations described above, I need to measure the fraction of college graduates employed in “noncollege” occupations. Thus, I need to classify all occupations where college graduates work into “college” and “noncollege” ones. In doing so I follow Gottschalk and Hansen’s (2003) approach based on the model presented in Section 3.1. 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 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, I 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.[18] 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 Czech Republic is significantly higher than in the U.S., I use a higher threshold (15%). Nevertheless, as presented in Section 6.4, 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 Czech labor market, I lower this threshold to 85%. Additionally, I classify occupations where more than 95% of workers have a high school diploma as “noncollege” occupations.

[18] 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.

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, I estimate the following wage equation:

$$\log w_{ik} = \beta_{0k} + \beta_{1k} \cdot exp_i + \beta_{2k} \cdot exp_i^2 + \beta_{3k} \cdot female_i + \phi_k \cdot coll_i + \varepsilon_{ik}, \quad (19)$$

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 dummy variable equal to 1 if worker has a college degree and 0 otherwise.[19] 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 $\phi_k$, 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 I use the Czech national employer survey, ISPV. This is a linked employee-employer data (LEED) gathered and processed according to the requirements of the Czech Ministry of Labor and the European Union. Information is collected from a sample of more than 3500 firms in the commercial sector that report wages and other information for about 1.3 million of workers. This dataset is a repeated cross-section; the data is collected on the firm-level and individual workers are not explicitly followed.

[19] 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.

The main advantage of the dataset used is its size. In order to apply the Gottschalk and Hansen (2003) methodology of classifying occupations, it is necessary to have no less than 100 observations of workers with high school or higher level of education in each occupation. In the ISPV dataset there are about 35,000 young college graduates, defined as individuals with at least a bachelor degree, below 35 years of age, and 65,000 young high school graduates, defined as individuals below 35 years of age[20] who have passed a maturity exam, for each of the years in the $2000-2008$ period. This is enough to carry out the analysis on the level of 3-digit occupations.

The variables reported in the dataset used include age, gender and education level of each employee. Moreover, one can find the characteristics of the firm in which an individual is employed (location, industry, size, ownership structure, etc.), the occupation that an individual performs and her monthly earnings together with the number of hours worked. The last two variables allow me to calculate the hourly wage which is defined as the average pay per hour during the first quarter of a year.

Occupations are coded in the ISPV dataset according to a local system which follows the International Standard Classification of Occupations (ISCO). For the purpose of this study, I use occupations defined on a 3-digit level. This is the precision used also by Gottschalk and Hansen (2003). Occupations defined by 3-digit codes are detailed enough to capture quite narrowly defined jobs performed there and at the same are wide enough to contain the number of workers allowing me to perform the estimations. Nevertheless, some occupations had to be merged in order to achieve a larger sample size, in which case the aggregation was kept the same for each year of the analysis.

District- and region-specific data on population and labor force structure are taken from the Czech Labor Force Survey (LFS). This survey is representative on the regional (NUTS-3) level. To get ditrict level information, 1991 and 2001 Census data are used. 2001 values are extrapolated to other years of the analysis using region-specific growth rates calculated from the LFS. Additionally, the district information on registered unemployment gathered by the

[20] Card and Lemieaux (2001) show that younger and older workers are not perfect substitutes. I work just with young workers to avoid this issue.

Czech Ministry of Labor is used to calculate gender- and employment-specific unemployment rates.

## 6.2 Cross-sectional estimation on the district level

This section presents the second stage estimates of the relationship between the relative number of college graduates in the population and their fraction working in "noncollege" occupations, as described by equation (18b), in cross-district dimension. As shown in Table 3, this analysis gives some evidence to believe that the productivity spillover fom high concentration of skills is strong enough create improved employment possibilities for college graduates in districts where their stock is relatively high. Columns 1-3 of Table 3 report OLS estimates of the relationship of interest, while columns 4-6 report the estimates of the same relationship when the relative stock of college graduates in a district has been instrumented by the share of college graduates in the district population as of the end of communism. Outlying districts are gradually removed from the analysis. Prague and Brno, the two major cities of the Czech Republic, are eliminated because they have incomparably large share of college graduates in the local population and a high concentration of businesses. Additionally, I remove districts characterized by high migration of college-educated citizens, as discussed in Section 2.

[Table 3 here]

Looking at Table 3, one can see 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 significantly negative 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 not to be that negative. 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. 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 negative with only 80% confidence. Nevertheless, it is not estimated to be positive, which would be the expected result when no spillover effects are present.[21] This gives us some evidence supporting the hypothesis that a larger number of college graduates attracts skill-intensive capital and in this way improves the situation of highly educated workers in the district labor market.

## 6.3 Estimation on the panel of districts

Let me now turn to the estimates of the relationship between the relative amount of college graduates in the population and their fraction working in "noncollege" occupations in cross- and within-district dimension. Table 4 presents the results of such estimation. Columns 1-3 include the OLS estimates of the relationship under consideration while columns 4-6 include the fixed-effect (FE) estimates, where time-constant district effects have been differenced out. Similarly as in the case of cross-district analysis, separate analyses were performed with exclusion of Prague and Brno as well as high migration districts.

[Table 4 here]

It is interesting to see that in the over-time dimension, the estimates of the relationship between the share of college graduates in the district population and the fraction of them working in "noncollege" occupations are positive even under OLS. The fixed effect estimates are even higher, as expected. This suggests that the supply effect is stronger than the spillover effect[22] and an increase in the relative stock of college graduates in the local labor market worsens their employment situation.

The opposing results of cross-sectional and over-time analysis might be interpreted in the following way. Districts with historically determined higher supply of college graduates have attracted skill-complementing capital and offer more employment possibilities in "college"

[21] Let me remind that, according to equation (16), $\delta_2 > 0$. Thus a non-positive estimate of $\theta_1 = \delta_1 + \delta_2$ implies that $\delta_1 < 0$, i.e. that the spillover effect exists.

[22] Movement along a downward sloping demand curve is larger in scale than the shift of this curve.

occupations. Thus, the situation of college graduates is better in these regions. Nevertheless, by stimulating an increase in the stock of college graduates from year to year, districts are not able to attract enough capital to compensate for the supply effect, and thus over time we observe a positive relationship between the share of college graduates in district population and the fraction of them working in "noncollege" occupations. These could be thought of as long run and short run effects. Positive spillovers from high concentration of college graduates are found to be significant only in the long-run context.

## 6.4 Robustness check

It could be argued that the results presented above are specific to the definition of "college" occupations. Let me remind that an occupation is defined to be "college" when the wage premium it pays to college graduates exceeds 15% or when the proportion of college graduates working there exceeds 85%. These thresholds have been chosen specifically to reflect the conditions of the Czech economy. To show that the results are not driven by the chosen thresholds, let me present the outcomes of analogous estimations performed using an alternative definition of a "college" occupation, i.e., with the wage premium threshold set at 10% and proportion threshold at 90%. These are the values used in previous research to distinguish between the "college" and "noncollege" occupations. As seen in the Tables 5-6, the use of an alternative definition leads to qualitatively the same results.

[Table 5 here]

[Table 6 here]

Additionally, I check whether the noisy character of district-level data does not influence the results of panel estimations. As explained in Section 6.1, district-level data for non-Census years are derived from the Czech Labor Force Survey (LFS) which is not representative on district level. Thus, I repeat panel estimation on regional level (a region aggregates 5 districts, on average), for which data derived from the LFS is more reliable. The relevant estimates are presented in Table 7. One can see that they are qualitatively the same as district-level regressions.

[Table 7 here]

The other robustness checks involved including different forms of $\frac{L_C}{L_N}$ in the regressions and repeating the analysis on a panel of firms subsample. Niether of these brought additional insight to the analysis.

# 7 Conclusion

Estimation of the fraction of college graduates working in "noncollege" occupations proposed by Gottschalk and Hansen (2003) and applied further to the U.S., Portugal, UK and the Czech Republic reveals a consistent pattern. In every country this measure has been decreasing over time despite a significant growth in the relative number of college educated workers in the labor market. This observation triggers a question about the source of such a state of being. Possible explanations that have been discussed in this paper are that (1) exogenous SBTC shifts the demand for college graduates so much that it more than matches shifts in the supply of them, or (2) a higher supply of college graduates stimulates SBTC, creating, in a way, a demand for itself.

These phenomena are, of course, not mutually exclusive. Most probably they both happen simultaneously. However, it is important from the policy point of view to find out how strong is the self stimulating effect as opposed to exogenous shifts in demand for college graduates. In this paper, I take a look at Czech district-level variation in the supply of college educated labor to find some evidence for an increased number of highly educated labor attracting skill-intensive industries and endogenously shifting the demand for skills, but only across districts. In the over-time dimension, I have not identified a strong enough spillover effect that would compensate for the demand effect allocating a larger fraction of college graduates to "noncollege" occupations when their relative number in the labor market increases. This finding has clear implications for further expansion of the Czech state-funded higher education system. It suggests 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). Nevertheless, in the short run the supply of college seats should be a response to the observed level of demand for skills (explanation 1).

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

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

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

[8] CSO, Czech Statistical Office (2008) http://www.czso.cz/eng/redakce.nsf/i/people_and_society_s1m

[9] Donald, S G., Lang K. (2007) Inference with difference in differences and other panel data,*The Review of Economics and Statistics*, 89(2): 221–233.

[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] 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.

[13] 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.

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

[15] 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.

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

[17] Jurajda S., Terrel K. (2009) Regional unemployment and human capital in transition countries Forthcomming in *Economisc of Transition*.

[18] Katz L., Autor D. (1999) Changes in the Wage Structure and Earnings Inequality, in: Ashenfelter O., Card D. (eds.), Handbook of Labor Economics, Amsterdam.

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

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

[21] 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.

[22] OECD (2008) Education at a Glance 2008: OECD Indicators, www.oecd.org/edu/eag2008

[23] Roback J. (1982) Wages, rents and the quality of life, *Journal of Political Economy*, 90(6): 1257-1278.

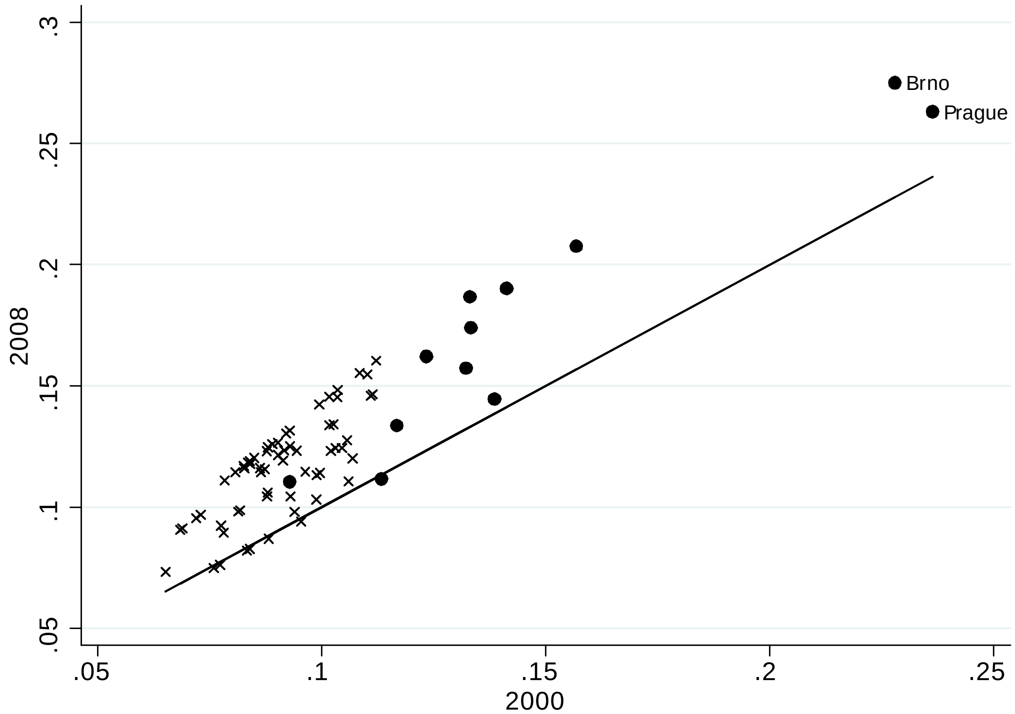
[24] Roback J. (1988) Wages, rents and amenities: Differences among workers and regions, *Economic Inquiry*, 26(1): 23-41.

[25] Wheeler C.H. (2001) Search, Sorting, and Urban Agglomeration, *Journal of Labor Economics*,19(4): 879-899.

# Appendix

## Figures

Figure 1: Changes in the fraction of college graduates in Czech districts' population between 2000 and 2008 together with a 45-degree line.




Note: Full circles denote districts which had a college by the end of communism, while crosses denote districts which did not have a college at that time. Growth rates are aggregated at region-level (NUTS-3) due to representative data availability.

Source: Own calculations using 2001 Census and the 2000-2008 Czech Labor Force Survey.

## Tables

Table 1: Summary statistics of the ISPV data

| Year | Total | Education | | Gender | |
|---|---|---|---|---|---|
| | | College | High school | Male | Female |
| 2000 | 123669 | 22% | 78% | 56% | 44% |
| 2001 | 134441 | 22% | 78% | 56% | 44% |
| 2002 | 134249 | 23% | 77% | 54% | 46% |
| 2003 | 138142 | 25% | 75% | 56% | 44% |
| 2004 | 164288 | 27% | 73% | 55% | 45% |
| 2005 | 173972 | 22% | 78% | 55% | 45% |
| 2006 | 185375 | 23% | 77% | 56% | 44% |
| 2007 | 220025 | 25% | 75% | 56% | 44% |
| 2008 | 231037 | 26% | 74% | 57% | 43% |

Note: The above table presents summary statistics of the sample of young workers, i.e., workers under 35 years of age.

Table 2: Changes in cohort-specific districts' population sizes (1991 - 2001)

| **Age in 1991** | **30 - 34** | **35 - 39** | **Age in 1991** | **30 - 34** | **35 - 39** |
|---|---|---|---|---|---|
| Benesov | 18% | 20% | Nymburk | 17% | 18% |
| Beroun | 9% | 16% | Olomouc | 11% | 12% |
| Blansko | 23% | 21% | Opava | 16% | 12% |
| Breclav | 11% | 12% | Ostrava-mesto | 4% | 5% |
| Bruntal | 5% | 9% | Pardubice | 8% | 4% |
| Ceska Lipa | 18% | 14% | Pelhrimov | 7% | 15% |
| Ceske Budejovice | 9% | 8% | Pisek | 1% | 6% |
| Cesky Krumlov | 12% | 12% | Plzen | 2% | 4% |
| Cheb | 19% | 16% | Plzen-jih | 16% | 16% |
| Chomutov | 14% | 10% | Plzen-sever | 15% | 18% |
| Chrudim | 10% | 9% | Prachatice | 4% | 10% |
| Decin | 16% | 19% | Prerov | 6% | 6% |
| Domazlice | 1% | 11% | Pribram | 1% | 4% |
| Frydek Mistek | 14% | 14% | Prostejov | 10% | 7% |
| Havlickuv Brod | 11% | 8% | Rakovnik | 21% | 12% |
| Hodonin | 11% | 10% | Rokycany | 5% | 20% |
| Hradec Kralove | 9% | 5% | Rychnov nad kneznou | 10% | 9% |
| Jablonec nad Nysou | 9% | 8% | Semily | 17% | 9% |
| Jicin | 12% | 13% | Sokolov | 16% | 9% |
| Jihlava | 10% | 9% | Strakonice | 1% | 7% |
| Jindrichuv Hradec | -11% | -20% | Sumperk | -16% | -13% |
| Karlovy Vary | 11% | 14% | Svitavy | 11% | 10% |
| Karvina | 10% | 8% | Tabor | 4% | 0% |
| Kladno | 15% | 12% | Tachov | 4% | 4% |
| Klatovy | 7% | 7% | Teplice | 18% | 17% |
| Kolin | 14% | 18% | Trebic | 8% | 11% |
| Kromeriz | 16% | 11% | Trutnov | 16% | 14% |
| Kutna Hora | 3% | 8% | Uherske Hradiste | 28% | 32% |
| Liberec | 6% | 9% | Usti nad Labem | 11% | 9% |
| Litomerice | 9% | 14% | Usti nad Orlici | 15% | 15% |
| Louny | 0% | 6% | Vsetin | 8% | 9% |
| Melnik | 10% | 9% | Vyskov | 11% | 6% |
| Mlada Boleslav | 15% | 14% | Zdar nad Sazavou | 11% | 6% |
| Most | 7% | 6% | Zlin | 7% | 8% |
| Nachod | 10% | 13% | Znojmo | 5% | 8% |
| Novy Jicin | 8% | 11% | **Country average** | **10%** | **10%** |

Table 3: Determinants of the share of college graduates in "noncollege" occupations across Czech districts in 2001

| | (1) OLS | (2) OLS | (3) OLS | (4) IV | (5) IV | (6) IV |
|---|---|---|---|---|---|---|
| $\ln\left(\frac{L_C}{L_N}\right)$ | -1.241** | -1.250** | -1.241** | -0.890 | -0.897 | -0.908 |
| (p-value) | (0.030) | (0.028) | (0.030) | (0.150) | (0.146) | (0.142) |
| Prague & Brno | Yes | No | No | Yes | No | No |
| High migration | Yes | Yes | No | Yes | Yes | No |
| Observations (distr. cells) | 71 | 69 | 67 | 71 | 69 | 67 |

Notes: The dependent variable is individual young college graduate's probability of working in a noncollege occupation (defined as paying a college premium higher than 15%). *CollShare* is the 2001 share of college graduates in a respective district's young population; as an IV for this variable, I use the share of college graduates in district population as of the end of communism (1991). Young workers are defined as being younger than 35. Columns (1) - (3) report OLS estimation results, while columns (4) - (6) report IV estimation results. P-values are in parentheses.

Table 4: Determinants of the share of college graduates in "noncollege" occupations across Czech districts over the 2000-2008 period

| | (1) OLS | (2) OLS | (3) OLS | (4) FE | (5) FE | (6) FE |
|---|---|---|---|---|---|---|
| $\ln\left(\frac{L_C}{L_N}\right)$ | 0.170** | 0.191*** | 0.214*** | 0.211** | 0.270** | 0.272** |
| (p-value) | (0.014) | (0.010) | (0.004) | (0.029) | (0.025) | (0.028) |
| Prague& Brno | Yes | No | No | Yes | No | No |
| High migration | Yes | Yes | No | Yes | Yes | No |
| Observations (distr.-year cells) | 639 | 621 | 603 | 639 | 621 | 603 |

Notes: The dependent variable is individual young college graduate's probability of working in a noncollege occupation (defined as paying a college premium higher than 15%). *CollShare* is the year-specific share of college graduates in a respective district's young population. Young workers are defined as being younger than 35. Columns (1) - (3) report OLS estimation results, while columns (4) - (6) report fixed-effect estimation results. P-values are in parentheses.

Table 5: Determinants of the share of college graduates in "noncollege" occupations across Czech districts in 2001

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | OLS | OLS | OLS | IV | IV | IV |
| $\ln\left(\frac{L_C}{L_N}\right)$ | -0.985* | -0.994* | -0.990* | -0.737 | -0.746 | -0.753 |
| (p-value) | (0.065) | (0.065) | (0.069) | (0.203) | (0.201) | (0.202) |
| Prague&Brno | Yes | No | No | Yes | No | No |
| High migration | Yes | Yes | No | Yes | Yes | No |
| Observations (distr. cells) | 71 | 69 | 67 | 71 | 69 | 67 |

Notes: The dependent variable is individual young college graduate's probability of working in a noncollege occupation (defined as paying a college premium higher than 10%). *CollShare* is the 2001 share of college graduates in a respective district's young population; as an IV for this variable, I use the share of college graduates in district population as of the end of communism (1991). Young workers are defined as being younger than 35. Columns (1) - (3) report OLS estimation results, while columns (4) - (6) report IV estimation results. P-values are in parentheses.

Table 6: Determinants of the share of college graduates in "noncollege" occupations across Czech districts over the 2000-2008 period

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | OLS | OLS | OLS | FE | FE | FE |
| $\ln\left(\frac{L_C}{L_N}\right)$ | 0.149** | 0.179*** | 0.201*** | 0.148* | 0.231** | 0.234** |
| (p-value) | (0.017) | (0.006) | (0.003) | (0.096) | (0.037) | (0.038) |
| Prague&Brno | Yes | No | No | Yes | No | No |
| High migration | Yes | Yes | No | Yes | Yes | No |
| Observations (distr.-year cells) | 639 | 621 | 603 | 639 | 621 | 603 |

Notes: The dependent variable is individual young college graduate's probability of working in a noncollege occupation (defined as paying a college premium higher than 10%). *CollShare* is the year-specific share of college graduates in a respective district's young population. Young workers are defined as being younger than 35. Columns (1) - (3) report OLS estimation results, while columns (4) - (6) report fixed-effect estimation results. P-values are in parentheses.

Table 7: Determinants of the share of college graduates in "noncollege" occupations across Czech regions over the 2000-2008 period

| | (1) OLS | (2) OLS | (4) FE | (5) FE |
|---|---|---|---|---|
| $\ln\left(\frac{L_C}{L_N}\right)$ | 0.050 | 0.024 | 0.178*** | 0.239*** |
| (p-value) | (0.524) | (0.766) | (0.081) | (0.043) |
| Prague | Yes | No | Yes | No |
| Observations (reg.-year cells) | 112 | 104 | 112 | 104 |

Notes: The dependent variable is individual young college graduate's probability of working in a noncollege occupation (defined as paying a college premium higher than 15%). *CollShare* is the year-specific share of college graduates in a respective region's young population. Young workers are defined as being younger than 35. Columns (1) - (2) report OLS estimation results, while columns (3) - (4) report fixed-effect estimation results. P-values are in parentheses.