Regional Studies

ISSN: 0034-3404 (Print) 1360-0591 (Online) Journal homepage: http://www.tandfonline.com/loi/cres20

Increasing social returns to human capital: evidence from Hungarian regions

László Czaller

REGIONAL STUDIES, 2017
VOL. 51, NO. 3, 467–477
http://dx.doi.org/10.1080/00343404.2015.1112898

Routledge
Taylor & Francis Group

Regional Studies Association

Increasing social returns to human capital: evidence from Hungarian regions

László Czaller

ABSTRACT

Increasing social returns to human capital: evidence from Hungarian regions. *Regional Studies*. Using individual-level data from 2002 to 2008, this paper estimates augmented Mincerian wage equations to analyze social returns to human capital in Hungary. The results show that geographically localized human capital externalities have a strong productivity effect on the wages of local workers, but the strength of this effect falls short of the private returns. A one-year increase in the average schooling of the local labour force has a 3% average external effect on the wages of local workers.

KEYWORDS

social return; human capital; education; productivity; wage; hungary

摘要

人力资本的社会报酬增加：来自匈牙利各区域的证据. *区域研究*。本文运用 2002 年至 2008 年的个人层级数据，评估增加的明瑟氏薪资方程，以分析匈牙利人力资本的社会报酬。研究结果显示，在地理上在地化的人力资本外部性，对于在地工作者的薪资有强大的生产力效应，但此般效应的强度却未符合私人报酬。在地劳动力平均增加一年的学校教育，对在地劳工的薪资产生了平均百分之三的外部效益。

关键词

社会报酬; 人力资本; 教育; 生产力; 薪资; 匈牙利

RÉSUMÉ

La croissance du rendement social du capital humain: des résultats provenant des régions hongroises. *Regional Studies*. Employant des données individuelles pour la période allant de 2002 à 2008, cet article estime des équations de salaire Mincer augmentées afin d'analyser les rendements sociaux du capital humain en Hongrie. Les résultats laissent voir que les effets externes du capital humain ont un effet productivité fort sur les salaires des travailleurs locaux en fonction de leur localisation géographique. Toujours est-il que l'ampleur de cet effet ne correspond pas aux rendements privés. Un relèvement d'un an de la scolarité moyenne de la main-d'oeuvre locale a un effet externe de 3% en moyenne sur les salaires des travailleurs locaux.

MOTS-CLÉS

rendement social; capital humain; éducation; productivité; salaire; hongrie

ZUSAMMENFASSUNG

Erhöhte Sozialerträge aus Humankapital: Belege aus ungarischen Regionen. *Regional Studies*. In diesem Beitrag werden anhand von Daten auf Individualebene im Zeitraum von 2002 bis 2008 erweiterte mincersche Lohngleichungen geschätzt, um die Sozialrendite von Humankapital in Ungarn zu analysieren. Aus den Ergebnissen geht hervor, dass geografisch lokalisierte Externalitäten des Humankapitals eine starke Produktivitätsauswirkung auf die Löhne lokaler Arbeitnehmer haben, wobei aber die Stärke dieser Auswirkung hinter der privaten Rendite zurückbleibt. Eine einjährige Verlängerung der durchschnittlichen Schulzeit lokaler Arbeitskräfte hat eine durchschnittliche externe Auswirkung in Höhe von 3% auf die Löhne lokaler Arbeitnehmer.

CONTACT

czallerlaszlo@hetfa.hu
Department of Regional Science, Eötvös Loránd University, Budapest, Hungary; and Center for Economic and Social Analysis, HÉTFA Research Institute, Budapest, Hungary.

SCHLÜSSELWÖRTER
sozialrendite; humankapital; bildung; produktivität; lohn; ungarn

RESUMEN
Aumento de la rentabilidad social del capital humano: evidencia de las regiones húngaras. *Regional Studies*. A partir de datos individuales de 2002 a 2008, en este artículo se calculan ecuaciones mincerianas aumentadas de salarios para analizar la rentabilidad social del capital humano en Hungría. Los resultados muestran que las externalidades del capital humano localizado geográficamente tienen un fuerte efecto de productividad en los salarios de los trabajadores locales, pero la fuerza de este efecto se queda rezagada en la rentabilidad privada. Un aumento de un año en escolaridad media de la fuerza laboral local tiene un efecto externo de un 3% de promedio en los salarios de los trabajadores locales.

PALABRAS CLAVES
rentabilidad social; capital humano; educación; productividad; salario; hungría

JEL D62, J24, R10
HISTORY Received January 2015; in revised form October 2015

INTRODUCTION

The subsidization of public and higher education is often justified by the argument that in the presence of positive human capital externalities, individuals systematically under-invest in their education while maximizing their utility. Economic agents are generally not aware of the social benefits (or costs) of their actions, or even if they were, these benefits would not be reflected in the market prices and it would be difficult to take stock of them. Another reason to subsidize education is the considerable role that human capital externalities play in promoting sustained economic growth (Lucas, 1988) and encouraging investment in physical capital (Acemoglu, 1996). Estimating the significance of social returns to human capital is therefore of particular importance.

Cross-country evidence (e.g., Benhabib & Spiegel, 1994; Klenow & Bils, 2000) shows that the average level of education explains a large part of subsequent total factor productivity (TFP) growth, suggesting an important role for human capital externalities in raising productivity. However, several problems arise at the national-level analysis. Measurement errors in cross-country education data and endogeneity issues caused by a large set of unobservable institutional factors raise serious concerns about the internal validity of cross-country studies (Krueger & Lindahl, 2001). The evaluation of social increasing returns to human capital at the local or regional level therefore seems more promising, as within-country institutional differences are relatively smaller and raw data sets on labour markets containing individual educational records are standardized. Moreover, the theoretical foundations of human capital externalities are closely related to the concept of regions or local labour markets. Skill acquisition through imitation, the transmission of tacit knowledge through contractual or informal non-market based channels and trust-building require direct interpersonal interactions and face-to-face contacts that facilitate the spatial clustering of individuals (Glaeser, 1999; Puga, 2010; Storper & Venables, 2004). It follows that external benefits of learning and knowledge spillovers decrease with distance because the costs of exchanging knowledge rise as distance grows (Arzaghi & Henderson, 2008; Jaffe, Trajtenberg, & Henderson, 1993; Rosenthal & Strange, 2008). Geographical boundedness (embodied, among others, in growing transactional costs and information loss) therefore suggests that analyzing human capital externalities at the local or regional level is reasonable as well.[1]

Glaeser, Scheinkman, and Shleifer (1995) reported a positive relationship between the human capital level of a city (usually measured as the share of college educated workers) and urban wages, but Rauch (1993) was the first to make an attempt to quantify econometrically the effects of human capital externalities using individual data. He estimated Mincerian wage equations augmented with a city-level average schooling term and found that average schooling has a positive and significant effect on the wages of local workers. Rauch argued that this effect on wages reflects productivity gains that are at least partly attributed to human capital externalities. Subsequent empirical studies, however, provided mixed results. Among others, Moretti (2004), Dalmazzo and Blasio (2007), and Iranzo and Peri (2009) confirmed the importance of human capital externalities, while Acemoglu and Angrist (2001) and Ciccone and Peri (2006) found no evidence in favour of them.[2]

The vast majority of the empirical work, however, focuses on the United States and just a few studies are concerned with other countries including the ones with transitional economies. This seems rather surprising because in these countries technology transfer through foreign direct investment (FDI) might be substantial sources of social increasing returns during and after the transitional period. Besides, in several Central and Eastern European (CEE) post-Communist countries, the educational system is highly centralized and the institutions are mostly financed by the public budget. According to the Organisation for Economic Co-operation and Development (OECD) (2012), in the year 2000 the public share of expenditures on education was above 90% in the Czech Republic, Poland, Slovak Republic and Hungary. Consequently, the magnitude of social returns should be of great importance in these countries when evaluating public investments on education. So far, however, only a few attempts have been made in these countries to analyze their importance (e.g., Muravjev, 2008).

To fill this gap and turn the attention to CEE countries, this paper continues the investigation on human capital externalities in Hungary by using an extensive dataset from repeated annual wage surveys of 2002–08. By using micro-data, it becomes possible to control for individual heterogeneity that otherwise would remain unobserved in regional- or country-level studies. Estimations are based on Mincerian wage equations augmented by additional firm and regional variables. To address endogeneity issues, the method of instrumental variables (IV) is used beside standard ordinary least squares (OLS) regressions. Historical census data on literacy rates and quarter of birth dummies are used to extract exogenous variation from local human capital and individual schooling. IV estimates on the whole sample of workers yield strong evidence on human capital externalities. Separate estimates for more and less educated workers suggest that these results are not likely to be driven by imperfect substitutability between workers with different educational backgrounds.

The remainder of the paper is organized as follows. The second section presents the model used in the empirical analysis and discusses identification issues related to its estimation. The third section introduces the empirical strategy and describes data. The estimation results and their implications are presented in the fourth section. The fifth section concludes.

EMPIRICAL MODEL AND IDENTIFICATION ISSUES

Empirical model

The most commonly used econometric framework setting up a relationship between the regional supply of human capital, as usually measured by the educational attainment of the local population and individual wages, takes the form of:

$$\log w_{i,a} = \log \alpha + \eta S_a + \varphi s_{i,a} \tag{1}$$

where $w_{i,a}$ represents the wage of individual i residing in region a; α is an intercept; S_a is average schooling in region a; and $s_{i,a}$ denotes years of schooling for individual i in region a. As shown by Acemoglu and Angrist (2000), equation (1) can be derived from at least two apparently different models. The first is a partial equilibrium model where externalities build in the production function in the form of technological increasing returns; the second is a random-search model where market interactions and asymmetric information in the labour market generate pecuniary externalities (Acemoglu, 1996). Since an empirical strategy based on equation (1) cannot distinguish between these types of externalities, thereinafter η is generally referred to as the strength of external returns to human capital irrespective of the exact source.

Endogeneity issues

Although different structural models justify the usage of the Mincerian approach for the evaluation of social returns, estimation based on equation (1) raises some identification issues. Since in reality there are several other confounding factors that affect individual productivity and wages, finding a positive and statistically significant coefficient for average schooling does not necessarily confirm the idea of social increasing returns to human capital.

The first problem is that parameter estimates based on Mincerian wage equations are potentially biased by endogeneity, which is partly related to spatial sorting. Individuals with certain types and amounts of unobserved abilities tend to move into regions where the returns to their abilities are likely to be higher. If these regions are those with high human capital endowments, the estimated coefficient of average schooling will also incorporate the effects of unobserved individual characteristics beside the social return to human capital. However, even if worker mobility is relatively low (as it is in Hungary), unobserved abilities may also affect the probability of getting a gainful job, especially in regions with a larger supply of skilled labour, which also causes spatial sorting with the same consequences. Furthermore, regions with higher human capital endowment may also have higher wages not only because of human capital externalities but also for a variety of other reasons. For example, average education is generally higher in large urban areas where local increasing returns may also arise from the sharing of facilities and specialized suppliers, labour pooling or better matching (Puga, 2010). Without controlling for these potential sources of increasing returns, the coefficient of average schooling will overstate the true magnitude of human capital externalities (Henderson, 2007). Nevertheless, more educated workers might be willing to pay more for housing and accept lower wages in order to enjoy cultural amenities unrelated to production (Falck, Fritsch, & Heblich, 2011), which in turn leads to a downward bias in the estimated coefficient on average schooling.

One straightforward way to overcome these issues and increase the likelihood that the estimate of η in equation (1) has a casual interpretation is the inclusion of a wide range of individual-, firm- and regional-level variables. To account for spatial sorting caused by unobserved worker heterogeneity, Duranton and Monastiriotis (2002) propose occupational dummies besides the conventional individual variables familiar from standard Mincerian equations. Dalmazzo and Blasio (2007) include firm-level variables and industrial dummies to control for spatial sorting patterns and also introduce a wide set of regional controls to increase the accuracy of the OLS estimates. Another solution is to estimate fixed-effects panel models, which, however, may still be biased by time-varying local demand shocks affecting wages and average schooling simultaneously. For example, during the restructuring stage of the market transition, Hungary experienced an upward shift in the demand for skilled labour which was partly due to the inward flow of FDI and the rapid diffusion of information and communication technologies (ICTs) (Commander & Köllö, 2008).

Particularly, the demand for skilled labour increased sharply in those regions where the level of human capital had been initially higher. Such region-specific demand shocks are difficult to observe, hence in order to identify correctly the magnitude of the social return, Acemoglu and Angrist (2001) proposed US state compulsory schooling laws and child labour laws as IV for average schooling. In the subsequent studies, other instruments were used such the presence of land grant colleges (Moretti, 2004) or the inflow of college-educated immigrants (Iranzo & Peri, 2009).

Another advantage of employing IV methods for identification is that they are also eligible to eliminate attenuation bias caused by measurement errors in the regressor of interest. Since this paper uses survey data to calculate aggregate schooling measures, this sort of bias is probably apparent and should be maintained.

Imperfect substitution between skill groups

With regard to the model outlined above, beside endogeneity there is a deeper conceptual problem that raises uncertainty about the interpretability of the results. If workers with different educational attainment are not perfect substitutes in production, then the regional supply of human capital might raise the productivity of unskilled workers and decrease the productivity and wages of skilled workers through simple factor substitution effects (Katz & Murphy, 1992). Hence, even in the absence of social returns, the aggregate supply of skills might increase average wages in the region. This is obviously problematic because in the Mincerian equations the effects of factor demand build in the estimated coefficient of the average schooling variable.

To deal with this issue, Moretti (2004) re-estimated the augmented version of equation (1) separately for each educational group, and compared the results. Finding that less educated workers benefit more from the aggregate supply of human capital than workers with higher educational attainment would confirm imperfect substitutability, however a positive and statistically significant effect on the wages of more educated workers would also suggest that positive externalities offset negative demand effects. Nevertheless, without being aware of the true magnitude of demand effects, this approach cannot exactly identify external effects and it only provides a rough approximation on their magnitudes (Ciccone & Peri, 2006).

IDENTIFICATION OF SOCIAL INCREASING RETURNS TO HUMAN CAPITAL USING HUNGARIAN DATA

By considering these issues, the estimation strategy is laid down as follows. In order to succeed in dealing with the problem of endogeneity, equation (1) is augmented by additional control variables. The empirical specification takes the following general form:

$$\log w_{i,a,t} = \log \alpha + \eta S_{a,t} + \varphi s_{i,a,t} + \mathbf{X}'_{i,a,t}\boldsymbol{\beta} + \mathbf{Z}'_{a,t}\boldsymbol{\delta} + \mu_a + \lambda_t + u_{i,a,t} \qquad (2)$$

where $\mathbf{X}'_{i,a,t}$ is a vector containing information on workers and their employers; $\mathbf{Z}_a$ is a vector of regional variables; $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are vectors of coefficients; and μ_a and λ_t denote regional and time-specific fixed effects respectively. Beside the conventional individual variables familiar from Mincerian wage equations such as gender, labour market experience and its square, vector $\mathbf{X}'_{i,a,t}$ also contains occupation dummies and additional employer characteristics including firm size, industry ownership and collective bargaining agreement coverage in order to capture unobserved abilities responsible for spatial sorting.

Vector $\mathbf{Z}_{a,t}$ consists of a number of wage determining factors at the regional level including employment density (number of employees per km^2) to isolate the effects of agglomeration economies unrelated to human capital, the ratio of bed places in commercial accommodation establishments to population, the ratio of crimes to population, and a dummy for the presence of United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage Sites as controls for local amenities. Besides, regional fixed effects at various spatial levels were also added to control for any remaining time-invariant local amenity endowments and structural characteristics. Another crucial control variable included is unemployment rate. As discussed by Blanchflower and Oswald (1995), bargaining and efficiency wage models deduce a negative relationship between local unemployment and individual wages (the so-called wage curve). Since unemployment rate is shown to be lower in regions where the supply of aggregate human capital is higher, the omission of unemployment rate from the model might therefore cause endogeneity bias.

Equation (2) is first estimated by OLS using pooled cross-sectional variation in regional characteristics. However, OLS estimates of social returns might be biased in the presence of time-varying region-specific demand shocks or other unobservables. In order to obviate this problem, regional-level average schooling is treated to be endogenous and estimated by two-stage least squares (2SLS). For this purpose, the ratio of literates among the population aged seven years and above in 1880 is used as an instrument for present-day average schooling. The validity of this instrument rests on the assumption that socio-economic factors that affected the spatial distribution of human capital in the late 19th century are not related to productivity and wages in the present day, apart from their remaining influences through the contemporary distribution of human capital. Late 19th-century literacy seems to meet this requirement because after the Austro-Hungarian Compromise in 1867 the spatial distribution of human capital in Hungary was just about to be reshaped by industrialization and the rapid expansion of the transport infrastructure.[3] Accordingly, the spatial distribution of human capital slightly after 1867 can be considered as the outcome of a predominantly agrarian small-scale feudal economic system. However, there might be other unobserved historical factors such as cultural norms, values or institutions that had a strong influence on literacy in the past and might also affect productivity to this day (Nunn, 2009). One way through which these cultural norms and values can be linked to education and long-term development is religious denomination. Recently, Becker and Woessmann (2009) provided empirical support in favour of Max Weber's famous hypothesis that states that the spread of the Protestant religion was instrumental in reducing illiteracy and facilitating industrial development in Europe. Since religious norms of behavior might transmit through generations, it is possible that they still have an influence on present-day productivity. The same might apply to class-specific attitudes and morality. Middle-class individuals practising occupations requiring skills, knowledge and marksmanship invested more time in learning, put more effort in working and developed entrepreneurial skills in the hope of upward social mobility and the well-being of their children (Doepke & Zilibotti, 2008). Bearing this in mind, additional proxies are added to the 2SLS specifications to ensure instrument exogeneity. The first variable that serves this purpose is the share of Protestant population in 1880, which proxies religious denomination; the second is a dummy variable that takes on the value of 1 if the sub-region is organized around a former royal free city. The idea behind this dichotomous variable is that the middle class consisting of artisans, merchants and shopkeepers was mainly located in these privileged cities at the beginning of the Hungarian industrial revolution. Intuitively, conditional on these long-term structural factors, 2SLS keeps only the variation in average schooling that is generated by 1880 literacy.

Another problem is that every instrument of regional schooling is necessarily correlated with individual schooling as well. Putting it differently, the exclusion restriction of the instrument might still be violated. To rule out this possible bias, quarter-of-birth dummies popularized by Angrist and Krueger (1991) were used to instrument individual years of schooling. According to the compulsory education laws operative between 1940 and 1990, children begin primary education after turning age six. Since the school year starts in September, students born in the last and earlier months of the year start first grade at an older age and reach the compulsory age of school attendance relatively earlier than those born in the second and third quarters of the year. Since season of birth is assumed to be randomly distributed over the population and supposedly uncorrelated with other wage-determining personal attributes, it seems an appropriate instrument in the Hungarian context.[4]

Finally, one more issue should be addressed here. Despite the choice of the estimation method, in models that rely on both individual and regional data the usual assumption that $u_{i,a,t}$ is identically and independently distributed does not hold, because regional innovations affecting local individuals drives to the clustering of the errors (Moulton 1986). This bias is corrected by using cluster-robust standard errors in every specification.

DATA AND DESCRIPTIVE OVERVIEW

The data needed for the analysis come from various sources. Individual data are from the Hungarian Wage Survey, which has been carried out by the National Employment Office each May since 1992.[5] The survey is only suitable to create pooled cross-sections; in the absence of any individual identifier it cannot be used to trace individuals over time and identify social returns through the time-varying wage outcomes of those who changed locations between two consecutive dates and those who stayed in the same local labour market.

Over the years the sampling of the survey has substantially changed; it only maintains the same structure since 2002, when it was developed to accomplish the requirements of the European Union Structure of Earning Surveys. Since 2002 the entire public sector, all firms employing more than 20 workers and a 20% random sample of firms employing fewer than 20 workers, have been included in the sample. The public sector and firms employing fewer than 50 employees provide data on all workers, while larger firms report a 10% random sample of their workers. This yields 100,000–220,000 observations per year for which personnel data on wages, hours of work, years of schooling, gender, age and occupation are available besides the information on the employer firms. After the reduction of the time span to seven years (2002–08),[6] and the exclusion of public servants, part-time workers and observations with missing data,[7] a dataset containing more than 900,000 observations (on average 137,000 observations per year) was retrieved.

Wages are defined as the logarithm of real gross average monthly wages expressed in 2002 HUF (Hungarian forints). Individual schooling is expressed in years; labour market experience is measured as age minus years of schooling minus six. To generate dummies for three broad occupational groups (manual, non-manual, managerial), detailed classes were aggregated along the hierarchy specified in Order No. 6/1992 of the Minister of Labor on the Inter-Sectoral Classification System of Employees. Firms are classified into six categories according to their number of employees (up to 10; from 11 to 20; from 21 to 50; from 51 to 300; from 301 to 1000; and 1001 or more). The foreign ownership dummy indicates whether at least 50% of the firm's capital is owned by foreign investors, and the collective bargaining agreement dummy indicates whether any of the employers are covered by an agreement. Finally, industry dummies correspond to the one-digit Nomenclature statistique des activités économiques dans la Communauté européenne (NACE) classification.

Although wage surveys provide information on job sites at the municipality level, the present analysis is rather conducted at the level of local administrative unit (LAU)-1 sub-regions. The main reason is that 174 LAU-1 regions provide a reasonable approximation to local labour markets in which the studied process is expected to operate. However, to the extent that LAU-1 regions lay within the commuting distance of workers, external effects might reach beyond the administrative boundaries. Neglecting spatial dependence caused by the inconsistencies of the arbitrarily defined borders and the true spatial extension of economic interactions might result in the overestimation of social returns if local wages and average schooling are correlated with the average years of schooling of neighboring regions. To test the importance of spatial dependence, a specification which includes the first-order spatial lag of average years of schooling is also estimated.

LAU-1-level variables stem from various sources. Average schooling and employment density are calculated directly from the micro-data using individual sample weights and the sub-regional classification of 2008. The source of unemployment rate and the amenity controls is the HCSO T-STAR database. Data on crime rates come from the Unified System of Criminal Statistics of the Investigative Authorities and of Public Prosecution. Finally, data on literacy rate and the share of Protestants are collected from 1880 Census records and transformed to be consistent with the 2008 LAU-1 classification.

Figure 1 depicts the spatial distribution of average wages and years of schooling in each sub-region for the period 2002–08. Panel A shows considerable spatial wage differences between rural and more urbanized areas. Sub-regions organized around or being in close proximity to bigger cities pay higher wages. While the metropolitan area of Budapest and highly urbanized sub-regions are in the upper quarter of the wage distribution, rural regions at the eastern border and less urbanized inner hinterlands are marked by low wages. On the other hand, the spatial distribution of human capital in panel B rather shows a zonal arrangement that follows the locations of urbanized areas and transport corridors. As a consequence, sub-regions at central and north-western Hungary are characterized by higher average educational attainment. Densely populated sub-regions and those areas that lie along the main highways stand out in their environments. Accordingly, spatial disparities of wages and schooling share some common patterns. Both variables are closely related to the level of urbanization; higher wages and schooling are usually found in large agglomerations, which draws attention to the possibility that raw correlations between wages and human capital are spurious due to the spatial sorting of workers or other sources of increasing returns related to urbanization.

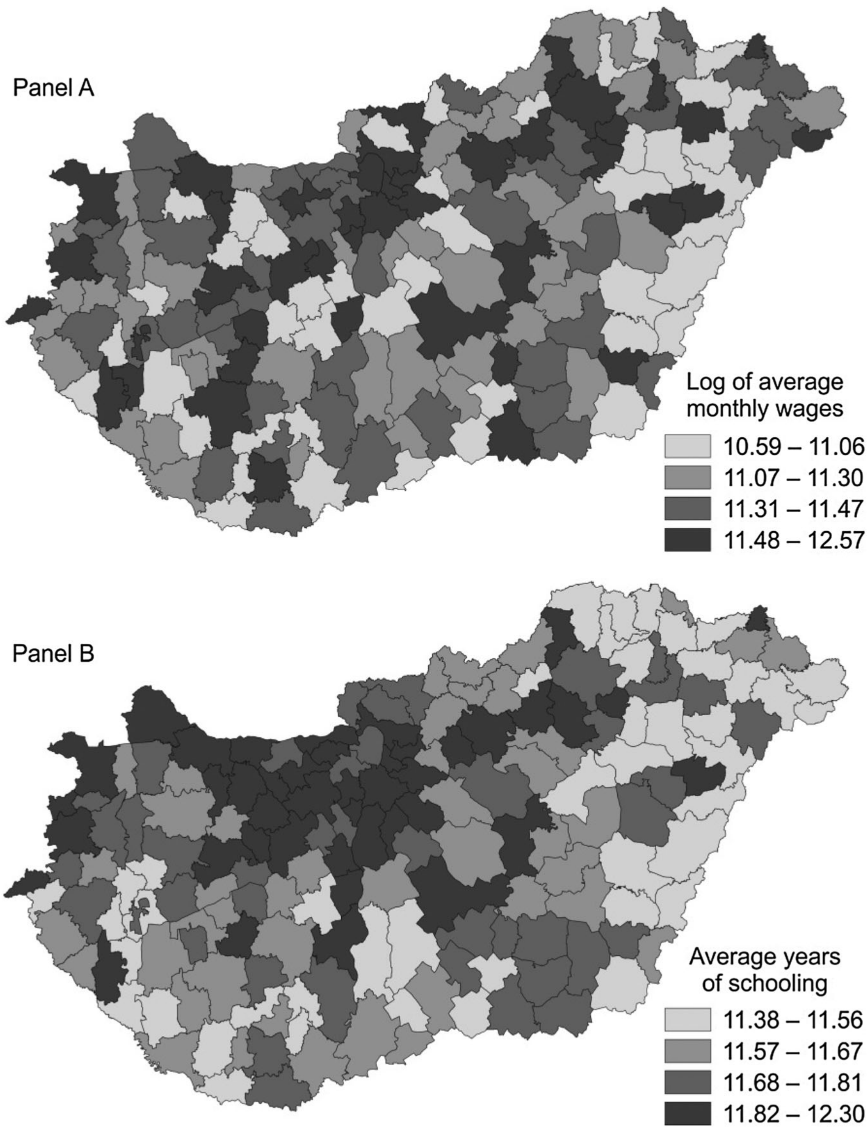

Figure 1. Log of average monthly wages and average years of schooling in local administrative unit (LAU)-1 sub-regions, 2002–08.

EMPIRICAL RESULTS

Basic results

The baseline results from OLS estimation of equation (2) are reported in Table 1. The first specification in column 1.1 only contains individual controls, six (of the seven) NUTS-2 dummies (Nomenclature des Unités Territoriales Statistiques), year dummies and a constant term. According to the coefficient on average years of schooling, the magnitude of social return is significant and slightly exceeds the private return to human capital. Nonetheless, this result is likely to be upward biased. Regressions in columns 1.2 and 1.3 add firm-level and regional controls to the model and report lower estimates. While firm-level characteristics seem to have a smaller impact on the coefficient of average schooling, it drops to half after the inclusion of regional controls (0.042). The results from the regression that includes a spatial lag for average schooling are presented in column 1.4. Parameter estimates do not change much compared with the specification in the previous column. The coefficient of the spatially lagged variable is nearly equal to the coefficient of local average years of schooling (about 0.04); however, due to its large standard error it remains insignificant. Column 1.5 includes LAU-1 fixed effects and drops the NUTS-2 and World Heritage dummies. The key coefficient is 0.025 and still remains significant at the 10% level. The results of the last column suggest that time-invariant regional characteristics upwardly bias OLS estimates without any fixed effects.

Endogenous average schooling

Although specifications reported in the last three columns of Table 1 are likely to provide more accurate estimates on the effects of average schooling than those without additional control variables, these parameter estimates might still be biased because of demand shocks, selective migration, unobserved regional characteristics (e.g., housing prices) or even measurement errors in the regional schooling variable. To preclude these problems, the literacy rate in 1880 is used as the instrument of average schooling. However, as noted above, any reasonable instrument of average schooling is also correlated with individual schooling. The correlation between individual years of schooling and late 19th-century literacy is 0.207 with $p = 0.000$, which raises some suspicion concerning the exogeneity of the instrument. Therefore, individual educational outcomes are also instrumented by quarter-of-birth dummies.

Table 2 reports the seasonal pattern in years of schooling by decades. The F-statistics in the bottom half of the table indicate that the joint effects of quarter of birth are significant. For every cohort born after 1940, the average completed years of schooling is 3–9% higher for those who born in the third quarter of the year than for individuals born in the first quarter. For the 1950s cohort, the difference between the completed years of schooling of those born in the first and second quarters of year is also significant. Except for the 1950s cohort, the effects of the fourth quarter are negative and significant. Table 2 shows that educational outcomes do vary by season of birth, however these within-year patterns are not uniform for the whole population: they vary substantially according to year of birth. To capture these differences in the relationship between schooling and quarter of birth across cohorts, a full set of interactions of birth years and quarter of birth are used as instruments where individual schooling is assumed to be endogenous.

The results of 2SLS estimates are reported in Table 3. In each specification the list of regional covariates is extended by proxies for persistent structural factors and norms to raise the probability of instrument exogeneity. Column 3.1 shows the just-identified case when only

Table 1. Effects of average years of schooling on wages (ordinary least squares (OLS) estimates).

	1.1	1.2	1.3	1.4	1.5
Average years of schooling	0.115***	0.087***	0.042***	0.043***	0.026*
	(0.018)	(0.014)	(0.017)	(0.018)	(0.013)
Individual schooling	0.105***	0.097***	0.097***	0.097***	0.097***
	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
Spatial lag of average schooling				0.039	
				(0.027)	
Individual controls	Yes	Yes	Yes	Yes	Yes
Firm controls		Yes	Yes	Yes	Yes
Industry dummies		Yes	Yes	Yes	Yes
Regional controls			Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Regional dummies	NUTS-2	NUTS-2	NUTS-2	NUTS-2	LAU-1
Adjusted R^2	0.364	0.518	0.519	0.519	0.524

Notes: Standard errors corrected for LAU-1 regional clustering are shown in parentheses. The number of observations is 958,167.
*Significance at the 10% level; ***significance at 1% level.

Table 2. Effects of quarter of birth on the educational outcome (ordinary least squares (OLS) estimates).

	1940–49	1950–59	1960–69	1970–79	1980–89
Quarter of birth (2nd quarter)	−0.053	0.046***	−0.010	0.027	−0.016
	(0.037)	(0.013)	(0.013)	(0.019)	(0.019)
Quarter of birth (3rd quarter)	0.090***	0.034**	0.029**	0.044***	0.070***
	(0.026)	(0.013)	(0.012)	(0.012)	(0.019)
Quarter of birth (4th quarter)	−0.058*	−0.033	−0.038***	−0.026**	−0.028*
	(0.026)	(0.023)	(0.013)	(0.012)	(0.012)
Joint test of quarter of birth dummies	8.325	14.858	12.130	5.643	39.574
(F-statistics, p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of observations	79 344	263 796	236 377	286 373	92 277

Notes: Robust standard errors are shown in parentheses.
*Significance at 10% level; **significance at 5% level; ***significance at 1% level.

average schooling is treated as endogenous. The critical value for the 10% maximal size distortion of 2SLS reported in Stock and Yogo (2005) is lower than the Cragg–Donald and the heteroskedasticity-robust Kleibergen–Paap F-statistics, which means that any bias arising from the weakness of the instrument is not probable. The estimated social return is 0.031 and invariably significant.

Column 3.2 treats both individual and average schooling as endogenous. Individual schooling is instrumented by 150 quarter-of-birth × year-of-birth interaction terms. Although the Sargan test of over-identifying restrictions confirms the exogeneity assumption of the instruments with $p = 0.142$, the drastically lowered Cragg–Donald and Kleibergen–Paap F-statistics suggest weak instruments. This might be because a large proportion of the interaction terms are uncorrelated with the schooling variable they are standing in for. Unfortunately, critical values of maximal size distortion are yet not available for such a large number of instruments, therefore it is hard to evaluate the magnitude of the bias caused by the weak instruments. Instead, in column 3.3 the same specification is re-estimated by limited information maximum likelihood (LIML), which is more robust on weak instruments than the 2SLS method. Compared with 2SLS, LIML provides

Table 3. Effects of average years of schooling on wages (two-stage least squares (2SLS) and limited information maximum likelihood (LIML) estimates).

	3.1		3.2		3.3	
Estimation method	2SLS		2SLS		LIML	
Average years of schooling	0.032***	(0.012)	0.029**	(0.012)	0.034**	(0.013)
Individual schooling	0.097***	(0.006)	0.070***	(0.010)	0.069***	(0.011)
Birth year dummies			Yes		Yes	
Individual controls	Yes		Yes		Yes	
Firm controls	Yes		Yes		Yes	
Industry dummies	Yes		Yes		Yes	
Year dummies	Yes		Yes		Yes	
Regional controls	Yes		Yes		Yes	
Regional dummies	NUTS-2		NUTS-2		NUTS-2	
Excluded instruments						
Average years of schooling	Literacy rate 1880		Literacy rate 1880		Literacy rate 1880	
Individual schooling			QoB × year of birth		QoB × year of birth	
Cragg–Donald F-statistic	4768.65		280.253		274.768	
Kleibergen–Paap F-statistic	22.251		8.207		7.791	
Sargan test of over-identifying restrictions			167.553	(0.142)	166.822	(0.151)

Notes: Standard errors corrected for local administrative unit (LAU)-1 regional clustering are in parentheses. The number of observations is 958,167. QoB, quarter of birth.
**Significance at 5% level; ***significance at 1% level.

very similar results. Parameter estimates and standard errors do not change in substantive ways, with the estimated coefficients of 0.029 and 0.034 regional schooling remaining significant at the 5% level. Due to the similar estimates, it seems that weak instruments only cause a minor problem.

Instrumenting individual schooling lowers the estimates of the private return to 0.070. However, in the case of average schooling, second-stage results do not show considerable differences. The coefficient of the variable of interest varies in the range of 0.029 and 0.034 depending on the specification, and remains significant at the 5% level in every case. On the whole, estimates in Table 3 suggest that an additional year of average schooling leads to a 3% increase in the wages of local individuals.

Testing for imperfect substitution

According to the results shown in Tables 1 and 3, the aggregate supply of human capital has a large positive effect on wages. However, as noted above, this finding might also be driven by imperfect substitution between workers with different educational backgrounds. To examine the relative importance of social returns to human capital and demand effects caused by imperfect substitution, equation (2) is re-estimated for two skill groups using the share of skilled workers as an alternative measure to aggregate human capital. The reason for the change is that a measure based on the share of the high-skilled is more informative about substitution effects than average years of schooling, which takes the whole skill distribution into account. For this, the sample is separated into two parts. Following Kertesi and Varga (2005), workers without high-school graduation corresponding to fewer than 12 years of schooling are considered as low-skilled workers, and those who have attained at least high-school graduation are considered as high-skilled workers. A similar classification is used by Dalmazzo and Blasio (2007) in the Italian case.

For both educational groups the coefficients of the share of high-skilled workers are positive and statistically significant; however, OLS and 2SLS provide opposite results. OLS results shown in Table 4 are in contrast to the arguments on imperfect substitution because they show that the relative supply of skilled workers has a larger effect on the wages of more educated workers irrespective of whether NUTS-2 or LAU-1 fixed effects are added to the model. On the contrary, when aggregate human capital is treated as endogenous, 2SLS results suggest that the argument on imperfect substitution seems to be valid because the external effect of education on low-skilled workers is somewhat larger than the effect estimated for their high-skilled peers.[8] For the subsample of more educated workers, 2SLS in column 4.6 provides lower estimates than OLS does in column 4.4, while 2SLS and the OLS specification that also includes LAU-1 fixed effects yield similar results. In the case of low-skilled workers, however, 2SLS estimates are somewhat larger than fixed-effect OLS estimates in column 4.2. The reason for the differing results is that 2SLS accounts for the wage effects of time-varying demand shocks, unobserved covariates and measurement errors, while OLS does not. Since skill upgrading during the economic transition suggests the importance of skill-biased demand shocks in the

Table 4. Skill-specific effects of the aggregate human capital on wages (ordinary least squares (OLS) and two-stage least squares (2SLS) estimates).

	4.1	4.2	4.3	4.6	4.4	4.6
Skill group		**Low-skilled workers**			**High-skilled workers**	
Estimation method	OLS	OLS	2SLS	OLS	OLS	2SLS
Share of high-skilled workers	0.228***	0.137**	0.164***	0.259***	0.153**	0.149**
	(0.042)	(0.062)	(0.034)	(0.059)	(0.069)	(0.062)
Individual schooling	0.065***	0.063***	0.065***	0.149***	0.148***	0.151***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regional controls	Yes	Yes	Yes	Yes	Yes	Yes
Regional dummies	NUTS-2	LAU-1	NUTS-2	NUTS-2	LAU-1	NUTS-2
Cragg–Donald *F*-statistic			9030.78			8104.52
Kleibergen–Paap *F*-statistic			44.080			23.797
Adjusted R^2	0.328	0.339	0.328	0.487	0.492	0.486
N	452,611			505,556		

Notes: Standard errors corrected for local administrative unit (LAU)-1 regional clustering are in parentheses.
**Significance at 5% level; ***significance at 1% level.

Hungarian case, the preferred estimation method is unequivocally 2SLS, which shows the sign of demand effects stemming from imperfect substitution but also confirms the existence of the external returns of human capital. As shown by the results in the last column, the effect of aggregate human capital on the wages of highly qualified workers is positive and significant at the 5% level, which means that positive externalities offset the negative effects of substitution. It is consistent with the results for the whole sample, suggesting that in Hungary the hypothesis of external returns to human capital is reasonable.

CONCLUSIONS

The purpose of this paper has been to provide further evidence in favour of human capital externalities. Mincerian wage equations augmented by an average schooling term and other controls were estimated in order to identify the social returns to human capital in Hungary. Although this methodology solves some measurement and endogeneity problems occurring in cross-country studies, it is only capable of identifying local effects of knowledge spillovers and pecuniary externalities.

According to the baseline estimates, social returns are about 2–4%. These results seem to be robust on the inclusion of LAU-1 fixed effects and the spatial lag of average schooling. More precise 2SLS and LIML estimates that exploit long lagged historical data on literacy are in a similar degree, regardless of whether or not individual schooling is treated as endogenous. Separate estimates for more and less educated workers confirm the hypothesis of imperfect substitutability between skill groups; however, 2SLS estimates show that the external effects of human capital exceed the negative effects of substitutability. According to the preferred 2SLS specifications in Table 3, the estimated social return is about 3–3.4%, which is smaller than the private return. For further comparison, social returns are about the one-sixth of the gender wage gap but have a larger impact on wages than does local unemployment.

The results are highly indicative of social returns to human capital, however further analysis is needed to understand where human capital externalities come from in the case of Hungary. Although social returns can be identified by using IV methods on a pooled cross-sectional dataset, the results do not say anything about whether external effects come from movers or stayers and explore the relative importance of selective migration in regional wage differences. This is obviously a weak point of the analysis.

From the policy point of view, however, the message of the results is straightforward. Private returns presumably underestimate the true economic value of human capital and education in Hungary. Therefore, when evaluating investments to education, it is crucial to consider not only the private but also the social returns to human capital. Moreover, geographically bounded external effects in the order of 3% justify the promotion of local skill formation, and, more importantly, the attraction and retention of skilled labour as relevant tools of regional policy.

ACKNOWLEDGEMENTS

The author thanks Bálint Herczeg, János Köllö, Klára Major and two anonymous referees for their helpful comments.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

NOTES

1. Beside knowledge spillovers and their positive effects on investment decisions, human capital also shapes voting behaviour and helps to reduce crime. Moreover, it might also have negative signalling effects (Spence, 1973).
2. For a systematic survey of the human capital externalities literature, see, for example, Lange and Topel (2006) or Heuermann, Halfdanasron, and Suedekum (2010).
3. Act XXXI of 1880 on local railways encouraged the connection of local railway companies to the state-owned railroad network. Consequently, during the following decades, the railroad network underwent a considerable expansion which in turn drastically reduced transport costs and facilitated migration.
4. Another possibility would have been to exploit changes in the regulations on the upper age limit between 1940 and 1990 and use them as instruments. During the 20th century these regulations changed several times, however they always remained state competence and never varied across regions. Preliminary calculations did not show any relationship between temporal changes of the laws on compulsory schooling and individual educational outcomes.
5. The database is maintained and updated by the Institute of Economics of Hungarian Academy of Sciences.
6. In 2008, the NACE classification was revised and the conversion between the old and new nomenclature would have implied a significant loss of data. Therefore, all available years after 2008 were excluded from the analysis.
7. Those who work fewer than 36 hours per week were considered as part-time workers.
8. Since date of birth primarily affects the distribution of schooling in the range of six to 12 years, using quarter of birth and birth year interactions as instruments is not appropriate for the subsample of workers bearing more than 12 years of schooling. Therefore, in the case of high-skilled workers, individual schooling is assumed to be exogenous. In the case of low-skilled workers, quarter-of-birth instruments would be appropriate, but according to the preliminary calculations, using instruments for individual schooling does not make any quantitative difference from the results of column 4.3.

REFERENCES

Acemoglu, D. (1996). A microfoundation for social increasing returns in human capital accumulation. *Quarterly Journal of Economics*, *111*, 779–804. doi:10.2307/2946672

Acemoglu, D., & Angrist, J. (2000). How large are human capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, *2000*, 9–59. doi:10.3386/w7444

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, *106*, 979–1014. doi:10.2307/2937954

Arzaghi, M., & Henderson, V. J. (2008). Networking off madison avenue. *Review of Economic Studies*, *75*, 1011–1038. doi:10.1111/j.1467–937X.2008.00499.x

Becker, S., & Woessmann, L. (2009). Was Weber wrong? A human capital theory of protestant economic history. *Quarterly Journal of Economics*, *124*, 531–596. doi:10.1162/qjec.2009.124.2.531

Benhabib, J., & Spiegel, M. M. (1994). The role of human capital in economic development evidence from aggregate cross–country data. *Journal of Monetary Economics*, *34*, 143–173. doi:10.1016/0304-3932(94)90047-7

Blanchflower, D. G., & Oswald, A. J. (1995). An introduction to the wage curve. *Journal of Economic Perspectives*, *9*, 153–167. doi:10.1257/jep.9.3.153

Ciccone, A., & Peri, G. (2006). Identifying human–capital externalities: Theory with applications. *Review of Economic Studies*, *73*, 381–412. doi:10.1111/j.1467-937X.2006.00380.x

Commander, S., & Köllö, J. (2008). The changing demand for skills evidence from the transition. *Economics of Transition*, *16*, 199–221. doi:10.1111/j.1468-0351.2008.00303.x

Dalmazzo, A., & Blasio, G. (2007). Social returns to education in Italian local labor markets. *Annals of Regional Science*, *41*, 51–69. doi:10.1007/s00168-006-0081-7

Doepke, M., & Zilibotti, F. (2008). Occupational choice and the spirit of capitalism. *Quarterly Journal of Economics*, *123*, 747–793. doi:10.1162/qjec.2008.123.2.747

Duranton, G., & Monastiriotis, V. (2002). Mind the gaps: The evolution of regional earnings inequalities in the U.K., 1982–1997. *Journal of Regional Science*, *42*, 219–256. doi:10.1111/1467-9787.00257

Falck, O., Fritsch, M., & Heblich, S. (2011). The phantom of the opera: Cultural amenities, human capital, and regional economic growth. *Labor Economics*, *18*, 755–766. doi:10.1016/j.labeco.2011.06.004

Glaeser, E. L. (1999). *Learning in cities*. NBER Working Papers No. 6271. Cambridge, MA: National Bureau of Economic Research (NBER). doi:10.3386/w6271

Glaeser, E. L., Scheinkman, J. A., & Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of Monetary Economics*, *36*, 117–143. doi:10.1016/0304-3932(95)01206-2

Henderson, V. J. (2007). Understanding knowledge spillovers. *Regional Science and Urban Economics*, *37*, 497–508. doi:10.1016/j.regsciurbeco.2006.11.010

Heuermann, D., Halfdanasron, B., & Suedekum, J. (2010). Human capital externalities and the urban wage premium: Two literatures and their interrelations. *Urban Studies*, *47*, 749–767. doi:10.1177/0042098009352363

Iranzo, S., & Peri, G. (2009). Schooling externalities, technology, and productivity: Theory and evidence from US States. *Review of Economics and Statistics*, *91*, 420–431. doi:10.1162/rest.91.2.420

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, *108*, 577–598. doi:10.2307/2118401

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics*, *107*, 35–78. doi:10.2307/2118323

Kertesi, G., & Varga, J. (2005). Foglalkoztatás és iskolázottság Magyarországon. *Közgazdasági Szemle*, *52*, 633–662.

Klenow, P. J., & Bils, M. (2000). Does schooling cause growth? *American Economic Review*, *90*, 1160–1183. doi:10.1257/aer.90.5.1160

Krueger, A. B., & Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature*, *39*, 1101–1136. doi:10.1257/jel.39.4.1101

Lange, F., & Topel, R. (2006). The social value of education and human capital. In E. Hanushek, & F. Welch (Eds.), *Handbook of the economics of education* (pp. 459–509). Amsterdam: North-Holland. doi:10.1016/S1574-0692(06)01008-7

Lucas, R. J. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*, 3–42. doi:10.1016/0304-3932(88)90168-7

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross–sectional data. *Journal of Econometrics*, *121*, 175–212. doi:10.1016/j.jeconom.2003.10.015

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics*, *32*, 385–397. doi:10.1016/0304-4076(86)90021-7

Muravjev, A. (2008). Human capital externalities evidence from the transition economy of Russia. *Economics of Transition*, *16*, 415–443. doi:10.1111/j.1468-0351.2008.00330.x

Nunn, N. (2009). The importance of history for economic development. *Annual Review of Economics*, *1*, 65–92. doi:0.1146/annurev.economics.050708.143336

Organisation for Economic Co-operation and Development (OECD). (2012). *Education at a Glance*. Paris: OECD.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, *50*, 203–219. doi:10.1111/j.1467-9787.2009.00657.x

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics*, *34*, 380–400. doi:10.1006/juec.1993.1042

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*, 373–389. doi:10.1016/j.jue.2008.02.006

Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics*, *87*, 355–374. doi:10.2307/1882010

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. Anderson (Ed.), *Identification and inference for econometric models* (pp. 80–108). New York: Cambridge University Press.

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*, 351–370. doi:10.1093/jnlecg/lbh027