The Geographical Reach of Human Capital Externalities

Johann Eppelsheimer*

February, 2017

Keywords: human capital externalities, geo-referenced data, smart cities, spillover effects, regional labor markets

JEL classification: D62; J24; J31; R10

Extended Abstract

Workers are not isolated beings. They interact with colleagues and learn from each other. Such human capital externalities are an old theme in regional economics (e.g. Marshall 1890; Lucas 1988). Knowledge spillovers are also seen as a main driving force behind technological change and productivity advancement (Acemoglu 1998). This is especially true if high-skilled workers, who play a crucial role in a knowledge society, are involved.

The literature investigating the impact of locally concentrated human capital on individual productivity is growing steadily (e.g. Rauch 1993; Acemoglu and Angrist 2001; Moretti 2004, Ciccone and Peri 2006, Heuermann 2011, Bratti and Leombruni 2014). I want to add to the literature by measuring the geographical reach of these spillover effects in an approach inspired by Rosenthal and Strange (2008). Rosenthal and Strange (2008) use US census data and measure the concentration of human capital within concentric rings around workers. The smallest of these concentric rings has a radius of five miles. Their main findings show that human capital externalities are present and that these spillover effects attenuate in space. In my study I use geo-referenced data from Germany that allows me to have an even more detailed look at the geographical reach of human capital externalities. More precisely I use registry data from the Institute of Employment Research (IAB), Nuremberg that contains information about workers and firms. Moreover the data contains the exact place of work. This allows me to measure the intensity of human capital within very narrow

*Institute for Employment Research Nuremberg (IAB), University of Regensburg (e-mail: johann.eppelsheimer@iab.de)

concentric rings (e.g. exactly at the workplace, 500 meters around the workplace, 1000 meters around the workplace, ...).

Furthermore to address differences in formal degrees between East and West Germany I use a task based skill definition to define workers as highly skilled (Brunow and Blien, 2015). The task based definitions also account for over- and under-education of high-skilled workers (Autor et al., 2003). The panel structure of the IAB data also enables me to control for several sources of unobserved heterogeneity through the inclusion of fixed effects.

The main branch of literature on human capital externalities tries to measure human capital externalities by estimating Mincerian type wage equations, there log-wages are explained by the regional share of high-skilled workers or average regional schooling. Even though many studies draw on large panel data sets and thus are able to control for unobserved heterogeneity through worker and region fixed effects endogeneity remains problematic. Scholars are concerned that human capital variables are endogenous due to region-specific and time-specific labor market shocks. If local labor market shocks affect wages and regional human capital simultaneously ordinary least squares estimates are biased.

Typically instrumental variable approaches are used to deal with this problem. Even if relevant instruments, that satisfy the exclusion restriction, are available instrumental variable approaches may have certain disadvantages. For instance many studies employ time constant instruments and thus discard the benefits of panel data. Depending on the instrument also the generalization of instrumental variable estimates can be limited because implications usually only apply to the group of compliers.

In a study on human capital externalities with German data Heuermann (2011) finds even larger wage effects of regional human capital compared to an ordinary least squares approach. He uses the number of public schools and students as instruments for the regional share of high-skilled workers. Therefore, overestimation of the spillover effects of human capital due to region-specific and time-specific labor market shocks might not be too problematic in Germany. However, as I cannot entirely rule out the endogeneity problem *a priori*, I aim to use several variables to control for regional labor market shocks that affect individual wages and the share of high-skilled workers simultaneously.

Besides region and time fixed effects I add variables that address regional labor market shocks caused by local and global sources. To address shocks from local sources, I first add the regional unemployment rate as a rough proxy of the current local labor market condition. To control for changes in the popularity of places I further include a proxy for amenities, namely the logarithmized number of hotel beds per capita. And to control for changes in the regional firm composition I add separate controls for openings and closures of small and large establishments to the model.

In order to address shocks stemming from global sources I expand the econometric model with three sets of interaction dummies. In the first set I interact year and industry fixed effects and thus take into account consequences of overall industry transformations on wages. In the second set I interact year and occupation fixed effects. This set of interactions absorbs shocks related to occupational supply and demand transitions. In the third set I intend to control shocks related to firm types and thus include interactions of year and firm size dummies. Hereby I aim to cover temporal effects induced by market trends that asymmetrically affect differently sized firms. This, for example, includes shocks that affect large internationally operating firms differently than small locally operating firms. Addressing all these possible sources of regional labor market shocks I intend to minimize the described endogneity problem.

The following equation summarizes my estimation strategy.

$$\log w_{i,t,s} = \mathbf{S}_{i,t}\boldsymbol{\theta}_s + \mathbf{X}_{i,t}\boldsymbol{\alpha}_s + \mathbf{Y}_t\boldsymbol{\beta}_s + \boldsymbol{\delta}_s + \mathbf{Z}_t\boldsymbol{\gamma}_s + \epsilon_{i,t,s}, \quad s \in \{\text{high-skilled}, \text{low-skilled}\}$$

In the model, worker type specific individual log wages respond to the share of high-skilled workers within a set of concentric rings of different radii ($\mathbf{S}_{i,t}$). The equation accounts for workers characteristics, such as education, age and tenure ($\mathbf{X}_{i,t}$), establishment and regional characteristics, such as firm size and population density ($\mathbf{Y}_t$) and a set of fixed effects, namely: worker, region, year, industry and occupation fixed effects ($\boldsymbol{\delta}_s$). Besides I control for potential regional labor market shocks from various sources ($\mathbf{Z}_t$).

I analyze register data provided by the IAB, Nuremberg. The data set is a random sample of all employees subject to social security in Germany. Self-employed workers and public servants are excluded. The data set gives information on wage, age, education and further personal characteristics. Information on wages is highly reliable because employers have to face legal sanctions in case of misreporting. I restrict the data to regular full-time workers and thus exclude part-time workers, apprentices, interns, trainees and marginal employed workers. Furthermore, I only consider workers between 18 and 64.

Wage data is top coded because of the contribution assessment ceiling in the German social security system. This affects not more than 9% of the observations.

I use a standard imputation method proposed by Gartner (2005) to correct the affected top coded records. Occupations are classified based on Schimpl-Neimanns (2003). The IAB also provides information on the size, the industry and the location of establishments. I use a 1-digit industry classification that is consistent over the observation period (Eberle et al., 2014; Statistisches Bundesamt, Wiesbaden, 2003). Regional information refers to the work place of individuals, not their place of residence.

Additionally, I add regional characteristics of the 402 German counties ("*Landkreise und kreisfreie Städte*") to the data set. The Federal Institute for Research on Building, Urban Affairs and Spatial Development supplies information on population density, amenities (hotel beds) and the local unemployment rate (INKAR).

I define high-skilled workers based on their occupational tasks and compute the share of these high-skilled workers among all workers in various concentric rings around each individual. The observation period is 2007 to 2009.

In a related current research project Möller and I show a link between the share of high-skilled workers in German counties and wages (Eppelsheimer and Moeller, 2017). Based on these results I also expect a strong relationship between the share of high-skilled workers in concentric rings around individuals and their wages. However the geographical reach and attenuation of human capital externalities remains open until the analysis of the geo-referenced data set is finished.

References

Acemoglu, D. (1998). Why do new technologies complement skills? directed technical change and wage inequality. *The Quarterly Journal of Economics 113*(4), 1055–1089.

Acemoglu, D. and J. Angrist (2001). How large are human-capital externalities? evidence from compulsory schooling laws. In B. S. Bernake and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2000*, Volume 15, pp. 9–74. MIT Press.

Autor, David, H., F. Levy, and J. Murnane, Richard (2003). The skill content of recent technological change: An empirical exploration. *The Qu 118*(4), 1279–1333.

Bratti, M. and R. Leombruni (2014). Local human capital externalities and wages at the firm level: Evidence from italian manufacturing. *Economics of Education Review 41*, 161–175.

Brunow, S. and U. Blien (2015). Agglomeration effects on labor productivity: An assessment with microdata. *Region 2*(1), 33–53.

Ciccone, A. and G. Peri (2006). Identifying human-capital externalities: Theory with an application to us cities. *The Review of Economic Studies 488*(73), 381–412.

Eberle, J., P. Jacobebbinghaus, J. Ludsteck, and J. Witter (2014). Generation of time-consistent industry codes in the face of classification changes. Institute of Employment Research, Nuremberg. FDZ-Methodenreport 05/2011.

Eppelsheimer, J. and J. Moeller (2017). Dynamic wage effects of brain gain and brain drain - analyzing changes in the regional concentration of high-skilled workers. work in progress.

Gartner, H. (2005). The imputation of wages above the contribution limit with the german iab employment sample. Institute of Employment Research, Nuremberg. FDZ-Methodenreport 02/2005.

Heuermann, D. (2011). Human capital externalities in western germany. *Spatial Economic Analysis 6*(2), 139–165.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*, 3–42.

Marshall, A. (1890). *Principles of Economics*. London: MacMillan.

Moretti, E. (2004). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics 121*, 175–212.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics 34*(3), 380–400.

Rosenthal, S. S. and W. C. Strange (2008). The attenuation of human capital spillovers. *Journal of Urban Economics 64*(2), 373–389.

Schimpl-Neimanns, B. (2003). Mikrodaten-tools: umsetzung der berufsklassifikation von blossfeld auf die mikrozensen 1973-1998. ZUMA, Mannheim. ZUMA-Methodenbericht 2003/10.

Statistisches Bundesamt, Wiesbaden (2003). German classification of economic activities (wz 93).