DISCUSSION PAPER SERIES

IZA DP No. 3919

Regional Measures of Human Capital in the European Union

Christian Dreger
Georg Erber
Daniela Glocker

January 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Regional Measures of Human Capital in the European Union

Christian Dreger
DIW Berlin and IZA

Georg Erber
DIW Berlin

Daniela Glocker
DIW Berlin and IZA

Discussion Paper No. 3919
January 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3919
January 2009

ABSTRACT

Regional Measures of Human Capital in the European Union*

The accumulation of the human capital stock plays a key role to explain the macroeconomic performance across regions. However, despite the strong theoretical support for this claim, empirical evidence has been not very convincing, probably because of the low quality of the data. This paper provides a robustness analysis of alternative measures of human capital available at the level of EU NUTS1 and NUTS2 regions. In addition to the univariate measures, composite indicators based on different construction principles are proposed. The analysis shows a significant impact of construction techniques on the quality of indicators. While composite indicators and labour income measures point to the same direction of impact, their correlation is not overwhelmingly high. Moreover, popular indicators should be applied with caution. Although schooling and human resources in science and technology explain some part of the regional human capital stock, they cannot explain the bulk of the experience.

JEL Classification: I20, O30, O40, O52

Keywords: human capital indicators, regional growth

Corresponding author:

Christian Dreger
DIW Berlin
Mohrenstr. 58
10117 Berlin
Germany
E-mail: cdreger@diw.de

* The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216813.

1 Introduction

The accumulation of intangible assets like the education of the labour force or the abilities to participate in the innovation process play a key role for economic growth in countries and regions. Investments in knowledge and education can generate substantial returns over the long run. Human capital accumulation is a cornerstone in models of endogeneous growth, see the seminal papers of Lucas (1988) and Romer (1990). Some authors have treated human capital as an input to the production process like any other factors. Its accumulation leads to increased capital deepening and a period of accelerated growth (Mankiw, Romer and Weil, 1992). Others like Aghion and Howitt (1992) have emphasized the critical role for the discovery and adaption of new ideas and innovations. According to that view, human capital is essential to transform ideas and innovations into new processes and products. Therefore, the Lisbon and Barcelona European councils have stressed the important role of R&D and innovation. One goal is to increase the investment in R&D to 3 percent of GDP.

The policy implications of distinguishing between the role of education as a factor of production and a factor that facilitates the diffusion of technologies are quite substantial. In the former, the utility from an increase in education is equal to its marginal product, which is proxied for example in terms of higher income or a higher probability to stay in the labour force. In the latter, the benefit is expressed in terms of a sum of its impact for all future output levels, since education raises total factor productivity growth and the speed of technology diffusion. Moreover, the growth record might depend to a larger extent on the stock of human capital, rather than on the changes, see Romer (1989) and Benhabib and Spiegel (1994). However, Krueger and Lindahl (2001) did not find strong evidence for the human capital stock. Instead, changes of the respective variables seem to be more important for income growth.

Despite the theoretical claim for the vital role of human capital to explain the process of economic growth, empirical evidence has been not overwhelming. Variables on educational attainment often appear to be insignificant or show even the wrong sign in cross section or panel regressions, where regional GDP per capita growth is explained by initial income and a number of additional factors, including human capital measures. See for example Pritchett (2001). Other researchers have emphasized that the role of human capital is largely overstated and stressed the reversed direction of causality (Bils and Klenow, 1999). Nevertheless, the empirical results may also be driven by the poor quality of the data, see Cohen and Soto (2001) and De la Fuente and Doménech (2006). Therefore, the construction of indicators to investigate the impact of human capital and to test conflicting hypotheses on its transmission channels to long run economic growth is of central relevance.

Because human capital is a multidimensional phenomenon, suitable proxies are not easy to find. Many researchers have focused on educational attainment, since this information is readily available. Typical measures include the years of schooling or the percentage of the labour force with secondary or tertiary education or rates of enrollment, see Barro and Lee (1993, 2000). However, these variables approximate only particular elements and neglect other aspects of human capital resources, like training on the job, specific knowledge or the previous working experience. As a consequence, they might blur the actual impact of human capital.

The construction of composite indicators can be an important step forward to overcome these deficiencies. They are able to handle a broader range of aspects and transform complex information into a unique measure. Hence, they may be easier to interpret than a bulk of univariate indicators. On the other hand, judgement is highly involved at several stages of the construction process. For example, the selection and weighting of the ingredients could have a crucial impact on the results. Thus, sensitivity analysis is required as a check for robustness.

This paper provides a comprehensive analysis of alternative human capital indicators available at the EU regional level. Regions are defined according to the NUTS1 and NUTS2 classifications. Examining the spatial dimension can offer new insights. Most striking, the amount of information is tremendously enlarged, while the evidence is less affected from omitted variable bias. In contrast, studies based on the country experience rely on a high number of observations only if very heterogeneous economies are included. The heterogeneity cannot be captured in a cross section and is proxied by fixed effects in a panel environment. But even the latter approach is not fully convincing, as structural differences across countries are hardly constant over longer time intervals. As the EU or at least the old and the new member states are more homogeneous geographical areas, the quality of the results should be enhanced. Furthermore, regional innovation clusters and areas of economic growth not necessarily linked to national borders can be explored within this framework.

In addition to univariate measures of human capital, composite indicators are discussed. To examine the robustness of the results, different aggregation methods are considered. The reliability of alternative indicators is investigated by using the Krueger and Lindahl (2001) approach. In addition, indicators based on wage regressions are presented, see Mulligan and Sala-i-Martin (1997) and Gershuny and Kun (2002). As an example, the earnings potential in a region is estimated. Due to data availability, these regressions are carried out for only for German NUTS1 regions.

The paper is organized as follows: Univariate measures of human capital are presented in section 2. To get an impression on the location of innovation areas, the spatial distribution of knowledge and education is also addressed. Section 3 discusses basic methodologies to construct composite indicators and benchmarks to evaluate their overall performance. In particular, the information content of the indicators to capture human capital resources can be assessed via reliability ratios. After introducing the theoretical concepts, multivariate indicators are constructed in the next two sections. In section 4, univariate indicators are aggregated to obtain the composite measures. As an alternative, indicators based on labour income are derived from wage regressions (section 5). Section 6 offers the conclusions.

2 Univariate indicators for human capital

The initial step is to examine suitable indicators to proxy the human capital stock at a regional level. The indicators are important on their own, but can also be exploited as ingredients for the composite measures. The primary source is the structural indicators database provided by Eurostat (http://ec.europa.eu/eurostat). While the analysis refers both to the NUTS1 and NUTS2 level, the broader concept has to be preferred in general. Although the NUTS2 classification can show a more disaggregated picture of the distribution of human capital, only a few variables are reported at this level. Since the univariate indicators serve as ingredients for the composite measures, they should have, at least approximately, the same quality across the regions in the sample. Regional education and science and technology indicators are all available for the NUTS1 classification. Nevertheless, large gaps can be observed even in this dataset. As a further drawback, the time series dimension is often rather short, covering only the last 5 or 10 years of experience. Since the analysis of growth processes requires long time spans, the indicators may be better interpreted as snapshots for the human capital stock at the regional level.

Schooling variables include the number of students at different stages of the education system, such as the pre-primary, primary, secondary and tertiary level. Regional science and technology indicators refer to R&D expenditures and personnel, human resources related to science and technology, and employment in technology and knowledge intensive manufacturing and services sectors. Most statistics are also reported for the gender dimension.

Appropriate human resources are important to determine whether regions are able to participate in the innovation process. Information is reported for different age categories. The latter is subdivided into education, occupation and core. For example, people in the core group have a tertiary education in science and technology and are employed in line with their education. Persons exclusively educated or occupated fulfil only one of these properties. Roughly 20 percent of people aged between 25 and 64 in the EU-25 have a tertiary educational attainment, and almost 30 percent are employed in respective activities. High skilled people work mainly in knowledge intensive services sectors and to a less extent in the manufacturing industries.

-Figure 1 about here-

Although the regional distribution of the univariate indicators might be broadly similar, it is far from being unique. Figure 1 illustrates this point by looking at two indicators, i.e. the number of workers in the core group and the number of scientists and researchers. Both measures are available even at the NUTS2 level. In order to eliminate the size of the region, they are expressed in relative terms, i.e. as a percentage of the total labour force or employment, respectively. In the Scandinavian countries, Belgium, the Netherlands, Western Germany, Switzerland and Austria more than 30 percent of the workforce are employed in the core group. Highly qualified jobs are concentrated in national capitals such as London, Paris or Budapest, since headquarters and government institutions are often located at these places. However, the picture changes substantially, if the analysis is focused on scientists and researchers. With shares above 2 percent of total employment, Oslo, Vienna and Budapest are on the leading edge of the sample. Overall, the correlation between the indicators is 0.53. The differences in the spatial distribution emphasize the fact that the locations of research centers and universities are not very closely linked to locations where the majority of high skilled people actually work. This also underpins the usefulness of aggregate measures to describe the human capital stock in a region.

To investigate whether the particular human capital component is related to the process of economic growth, Barro type growth regressions can be used as a workhorse, see for example Sianesi and Van Reenen (2003). This approach can also test the suitability of univariate indicators in the composite index. Income per capita growth is regressed on several factors, including initial per capita income and particular human capital measures. However, as the regressions investigate the relationship between education inputs and economic outputs without looking at the process linking them, this approach should be applied with caution. The results may suffer from omitted variable bias and reversed causality, see Bils and Klenow (1999). Among others, education can respond to the anticipated rate of income growth.

Therefore, a two step regression procedure is involved. This approach estimates the relation between higher human capital investments and economic performance through a bridging indicator. The latter represents the concrete transmission channel of the human capital impact. For example, education spending contributes to the training of a skilled labour force in the first stage. In the second stage, the induced increase in skills is expected to improve the economic performance, as measured by higher productivity and income growth.

For illustration purposes, the impact of scientists and researchers on economic growth is explored using the two step procedure. In particular, more scientists trigger an increase in the number of people working in high quality jobs. This in turn should lead to higher growth of income per capita. The results are shown in table 1. As the variables need to be known over the same cross sections, the regressions are based on 185 NUTS2 regions, including 7 Norwegian areas.

-Table 1 about here-

All coefficients are well signed. In the first step, a positive relationship can be detected between core workers and scientists. In particular, an increase in scientists raises high skilled jobs to a larger extent. The fitted values from the former regression have a positive impact on growth in the second step. The negative sign of initial income reflects convergence of per capita income. Regional convergence takes place with a rate of 1.3 percent per annum.

3 Constructing composite indicators

As human capital has many facets, univariate indicators are not sufficient to describe the entire phenomenon. For example, the years of schooling is an important ingredient, but can be a biased estimate of the total stock of knowledge. If schools adapt to new technological situations only with some delay, schooling might increase while human capital might not. Working experience is not considered at all, which is a serious drawback in periods of fast technological change. Hence, a composite indicator could be favoured. It transforms various aspects into a unique measure and might be easier to interpret than its ingredients. However, different aggregation methods can blur the results. Thus, sensitivity analysis is indispensable to examine the robustness of the aggregate. An extensive discussion of these issues has been provided by Nardo, Saisansa, Saltelli and Tarantola, Hoffman and Giovannini (2005).

Apart from missing values problems, the construction process can be described as a three step procedure. First, the ingredients of the overall index have to be selected. The quality of the aggregate depends on the quality of the underlying series, where selection criteria like relevance, analytical soundness and accessibility are involved. The ingredients have to capture the different dimensions of human capital, such as schooling, working experience, or the use of key technologies. Second, the univariate measures have to be transformed into a same scale. For example, ratios can be used instead of the original variables, where the indicators are divided by a suitable benchmark like the EU average. Standardized scores can also be employed, where each measure is replaced by the difference between its observation from the average and divided by the standard error. The empirical moments in the standardization exercise refer to the regional distribution of the respective variable.

The third step is most critical and devoted to the weighting of the transformed variables in the composite index. From the huge set of possible techniques, two often used strategies are discussed to obtain some clues on the dispersion of the results. In the first variant, the weights of the individual series are restricted to be equal. The composite indicator coincides with the arithmetic average of its ingredients. Alternatively, the weights are determined by factor analysis. In fact, the composite indicator is defined to be the first common common component of the univariate variables. It arises as a linear combination of the latter, with weights equal to the correlation coefficients between the single variables and their aggregate. The first common component represents the maximum contribution to the total variance of the ingredients.

As a different construction principle, a composite indicator might also be based on potential labour income within a region and arise from wage regressions, see Mulligan and Sala-i-Martin (1997) and Gershuny and Kun (2006). Information on the microeconomic level like education, the employment record, and socio economic characteristics like gender and martial status can be used to explain wages. The coefficients from the equation are then used to predict the earings potential of the entire population. However, the results are subject to a sample selection bias, as wages are only observed for people who are actually in work. The censoring problem can be addressed by estimating a two-step Heckman selection model (Heckman 1979).

Krueger and Lindahl (2001) have proposed a procedure to investigate the information content of univariate or multivariate measures of human capital. Let H be the true stock of human capital and that $P_1=H+\varepsilon_1$ a noisy estimator for this variable. The measurement error ε_1 has white noise properties, i.e. zero mean, constant variance, no autocorrelation and is uncorrelated with H. The information content is defined to be the ratio of the signal to the signal plus measurement error. The reliability ratio

(1) $$r_1 = \operatorname{var} H / \operatorname{var} P_1 = \operatorname{var} H / (\operatorname{var} H + \operatorname{var} \varepsilon_1)$$

is bounded to the unit interval, where larger values represent a higher information content. As the true stock of human capital is unknown, the ratio (1) cannot be computed. This would require a second imperfect measure $P_2=H+\varepsilon_2$, where the measurement error is also white noise. Given that ε_1 and ε_2 are uncorrelated, the covariance between P_1 and P_2 can be used to approximate the variance of H. Thus, the reliability ratio for the first indicator can be estimated by

$$(2) \qquad \hat{r}_1 = \operatorname{cov}(P_1, P_2) / \operatorname{var} P_1$$

that is, by means of the slope coefficient of an OLS regression of P_2 on P_1. In principle, this regression gives an idea how well P_1 is able to explain the true human capital stock because the measurement error in the dependent variable (P_2) is expected to be absorbed by the usual regression disturbance without any biases. It should be emphasized that the measure (2) displays useful information only if P_1 and P_2 are already reliable measures, i.e. that they need to be unbiased and consistent. Deviations from the true human capital stock are supposed to be random. Systematic patterns in measurement errors can invalidate the whole concept. These assumptions can be relaxed to some extent (De la Fuente and Doménech, 2006).

4 Multivariate indicators of human capital

In principle, multivariate indicators arise as aggregates from the univariate series. Overall, 40 variables describing schooling and science and technology activities have been collected at the NUTS1 level. Although this is an encouraging high number, it is quite important to note that only partial information is available even at the level of broader regions. For example, if the analysis is restricted to cases where all variables are observed, more than 75 percent of the cross section have to be dropped. Thus the analysis has to be based on a subset of variables. Specifically, multivariate indicators can be based on certain subsets of the univariate ingredients. In particular, 16 series describing the level of schooling, human resources in science and technology and expenditures for research and development are available for 64 out of 97 regions, see table 2 for a list of regions and variables.

-*Table 2 about here*-

A multivariate indicator is constructed by applying two techniques. In the first variant, univariate measures are transformed to standardized scores and then aggregated with equal weights. Alternatively, the weights are given by the factor loadings obtained from a principal component analysis conducted at the EU level for the 2002-2004 period. The multivariate index is equal to the first common component of the underlying variables. The first component is able to represent 92 percent of the total variance of the univariate series.

-*Figure 2 about here*-

The EU weighting scheme is then applied to calculate the multivariate index at the regional level, see figure 2. As the time series are largely incomplete, the regional analysis provides a snapshot for the year 2003, where most information is available. But this is not a severe limitation, as the stock of human capital changes only gradually over time. Since the weights determined by principal component analysis turn out to be very close to those obtained by the equal weighting approach, the respective indicators are quite similar. In fact, their correlation is roughly 1. Therefore, figure 2 shows only the results of the factor approach. High factor scores can be observed in particular in the Western part of Germany, France, Italy, and the UK. In addition, the Hungarian regions have high levels in the multivariate indicator.

The composite index can be employed to examine the usefulness of popular indicators of human capital. In particular, it is seen as a proxy for the true human capital stock, as the aggregate covers different aspects of the phenomenon. In this setup, the quality of series like schooling, human ressources in science and technology, and R&D expenditures may be investigated by looking at their reliability ratios. The latter arise as the slope parameters from a regression of the composite indicator on the respective variables, see table 3 for the regression results.

-Table 3 about here-

All univariate measures are able to explain a substantial part of the human capital stock. However, their reliability ratios range only between 0.1 and 0.2, implying that the bulk of the variable is not captured by the indicators. Schooling (all students) and human ressources in science and technlogy outperform R&D expenditures in a region, as the reliability ratios are doubled.

5 Labour income measures of human capital

Labour based income measures of human capital are constructed using microeconomic datasets. For illustration, the following analysis refers to the 2005 wave of the German Socioeconomic Panel (GSOEP). The GSOEP is a unique household panel dataset conducted at DIW Berlin and covers a wide range of social and economic variables. Among others, it provides detailed information on working status, labour income, schooling and other education, work experience and various sociodemographic characteristics like marital status or household size.

Empirical evidence is based on 15,829 individuals, aged between 18 and 65 and part of the labor force. 52 percent are women. The hourly wage is observed only if someone is actually in work. However, inference should be conducted for the entire population. Therefore, the two-step Heckman selection procedure is applied to overcome the resulting censoring problem (Heckman 1979). In fact, the decision to work can be captured by a binary choice model,

$$(3) \qquad z_i^* = \alpha' w_i + u_i$$

where z_i^* is the underlying unobserved latent variable (i.e. the propensity to work), α is a vector of parameters, w_i the vector of variables explaining the decision to work and u_i a white noise error term. An individual i works whenever the latent variable exceeds a threshold ($z_i^*>0$). Hence, wages are determined as

$$(4) \qquad y_i = \beta' x_i + \varepsilon_i \quad , \quad y_i = \begin{cases} 0, & z_i^* \le 0 \\ y_i^*, z_i^* > 0 \end{cases}$$

where y_i is observed wage, y_i^* is corresponding wage for the entire population that can be revealed if someone works, x_i the vector of variables explaining the wage and ε_i the idiosyncratic error. The errors in (3) and (4) are jointly distributed as normal and can be contemporaneously correlated.

The first step refers to the probability to work, given the individual and household characteristics. The probability to participate (P_i)

(5) $$P_i(z_i^* > 0) = P_i(\alpha' w_i + u_i > 0) = P_i(\alpha' w_i \leq u_i) = \Phi(\alpha' w_i)$$

can be explained in a probit model, where $\Phi(z)$ is the cumulative distribution function of the standard normal. For each individual, the inverse Mills ratio

(6) $$\hat{\lambda}_i = \varphi(\hat{\alpha}' w_i) / \Phi(\hat{\alpha}' w_i)$$

is calculated, where $\varphi(z)$ denotes the probability density function of the standard normal. The inverse Mills ratio is used to control for the sample selection bias. Besides the individual and household characteristics the wage equation includes the estimated value of the inverse Mills ratio, i.e.

(7) $$\log(y_i) = \beta' x_i + \theta \hat{\lambda}_i + v_i \,.$$

Note that the sign of the coefficient of the inverse Mills ratio can provide useful information, as it indicates the correlation between the unobservables in the participation (5) and outcome equation (7). It shows how the wage affects the probability to work. In this sense, the standard *t*-test of the null hypothesis θ=0 can therefore be interpreted as a test of no selection bias.

The model is estimated with the same explanatory variables for the participation and outcome equation. As usual, the identification hinges on the non linearity of the inverse Mills ratio. However, the problem with such a model without further restrictions is that it may result in substantial collinearity between the predicted inverse Mills ratio and the remaining covariates in the outcome equation. Hence, exclusion restrictions are used in the subsequent analysis, and they refer to household characteristics.

Estimation is done separately for women and men, because of their heterogeneity with respect to participation and wages. According to Mincer (1965), the dependent variable is the log of the hourly wage. Regressors for the participation equation (5) are schooling and other education, years in unemployment and additionally for women years in part-time employment. Since the age of the individual enters the specification, these measures might be interpreted as deviations from the overall time spent in full-time employment, which serves as the base category. Since all of these variables can determine wages as well, they are also used for equation (7). Household features like the number of children and marital status are employed as exclusion restrictions for the participation equation. In addition to the individual characteristics, firm size and region (federal state) are also included. See table 4 for the results.

-Table 4 about here-

All the parameters are well signed. For example, higher education raises wages. Compared to a person without any qualification (i.e. only elementary schooling, but no job training certification) the wage of an individual with a high education (university degree) is about 50 (41) percentage points higher for women (men). Deviations from full-time employment have a negative impact. An additional year in unemployment rather than in full time employment reduces the hourly wage for both women and men. A similar effect is observed for women in part-time employment. Furthermore, the hourly wage tends to be lower in the Eastern states.

Using the wage equation, the uncensored expected value for the underlying wage $E(y^*)$ can be inferred. It is obtained as the predicted average of the dependent variable for the entire sample. Furthermore, the subgroup of the labour force without job qualification is considered separately. As a final step, the uncensored wage is multiplied with monthly hours worked. By using individual expansion factors, average monthly wages are calculated for each federal state. This reflects the regional earnings potential, see table 5 for the results.

-Table 5 about here-

Average income can be also estimated for low qualified workers, see the lower part of table 5. The differences between averaged income per person and averaged income per low qualified worker can be interpreted as a skill premium. Furthermore, high skill premia can be seen as an indication for excess demand of human capital, implying that the available ressources are too low. In fact, the correlation between the composite indicator based on the factor model and the skill premia is negative for the German regions, -0.29. While this coefficient has the expected sign, it is not overwhelmingly high in absolute value. This also points to an impact of the construction principles on human capital indicators.

6 Policy implications

The accumulation of the human capital stock plays a key role for the macroeconomic performance across regions. Despite the strong theoretical support for this claim, the empirical evidence has been not very convincing, probably because of the low quality of the data. This paper makes progress in providing a robustness analysis of alternative measures of human capital available at the level of EU NUTS1 and NUTS2 regions.

Human capital indicators available focus often on particular aspects of the overall phenomenon. For example, the years of schooling is an important ingredient, but can be a biased estimate of the total stock of knowledge. If schools adapt to new technological situations only with some delay, schooling might increase while human capital might not. Working experience is not considered at all in the schooling variable. This is a drawback in periods of fast technological change.

To overcome these deficits, new composite indicators are constructed. They transform various aspects of human capital into a unique measure that might be easier to interpret than the ingredients. As different aggregation methods can blur the results, a sensitivity analysis is required to examine the robustness of the aggregate. Therefore, composite indicators based on different construction principles are proposed. They rely on aggregation of individual facets of the human capital stock. In addition, a labour income based measure is presented to assess the skill component of the earnings potenial in a region. Because of data availability, the latter analysis is carried out at the level of German NUTS1 regions.

The analysis shows a significant impact of construction techniques on the quality of indicators. While composite indicators and labour income measures point to the same direction of impact, their correlation is not overwhelmingly high. Furthermore, the analysis documents that several popular indicators should be applied with caution. Variables like schooling, human ressources in science and technology or R&D expenditures are able to explain some part of the regional human capital stock. However, they cannot capture the bulk of the experience, implying that the empirical estimates for the true human capital impact might be biased. In this sense, composite indicators are superior. Overall, the analysis would certainly benefit from higher data quality. Strong effort is indispensable to fill the gaps in the existing databases.

References

Aghion, P., Howitt, P. (1992): A model of growth through creative destruction, Econometrica 60, 323-351.

Barro, R., Lee, J.W. (1993): International comparisons of educational attainment, Journal of Monetary Economics 32, 363-394.

Barro, R., Lee, J.W. (2000): International data on educational attainment: Updates and implications, CID Working Paper 42, Harvard University.

Bassanini, A., Scarpetta, S. (2001): Does human capital matter for growth in OECD countries? Evidence from pooled mean-group estimates, OECD Economics Department, Working Paper 282.

Benhabib, J., Spiegel, M.M. (1994): The role of human capital in economic development: Evidence from cross-country data, Journal of Monetary Economics 34, 143-173.

Bils, M., Klenow, P. (1999): Does schooling caused growth?, American Economic Review 90, 1160-1183.

Cohen, D., Soto, M. (2001): Growth and human capital: Good data, good results, CEPR Discussion Paper 3025.

De la Fuente, A., Doménech, R. (2006): Human capital in growth regressions: How much difference does data quality make?, Journal of the European Economic Association 4, 1-36.

Gershuny, J., Kun, M.Y. (2006): Wage earning potential from BHPS data, University of Essex, ISER Working Paper 2006-03

Krueger, A.B., Lindahl, M. (2001): Education for growth: Why and for whom?, Journal of Economic Literature 39, 1101-1136.

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

Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. The Quarterly Journal of Economics 107, 407-437.

Mulligan, C.B., Sala-i-Martin, X. (1997): A labor income-based measure of the value of human capital: An application to the states of the United States, Japan and the World Economy 9, 159-191.

Mulligan, C.B., Sala-i-Martin, X. (2000): Measuring aggregate human capital, Journal of Economic Growth 5, 215-252.

Nardo, M., Saisana, M., Saltelli, A.,, Tarantola, S. (EC/JRC), Hoffman, A., Giovannini, E. (OECD) (2005): Handbook on constructing composite indicators: Methodology and user guide, OECD Statistics Working Paper.

Pritchett, L. (2001): Where has all the education gone?, World Bank Review 15, 367-391.

Romer, P. (1989): Human capital and growth: Theory and evidence, NBER Working Paper 3173.

Romer, P. (1990): Endogeneous technological change, Journal of Political Economy 98, 71-102.

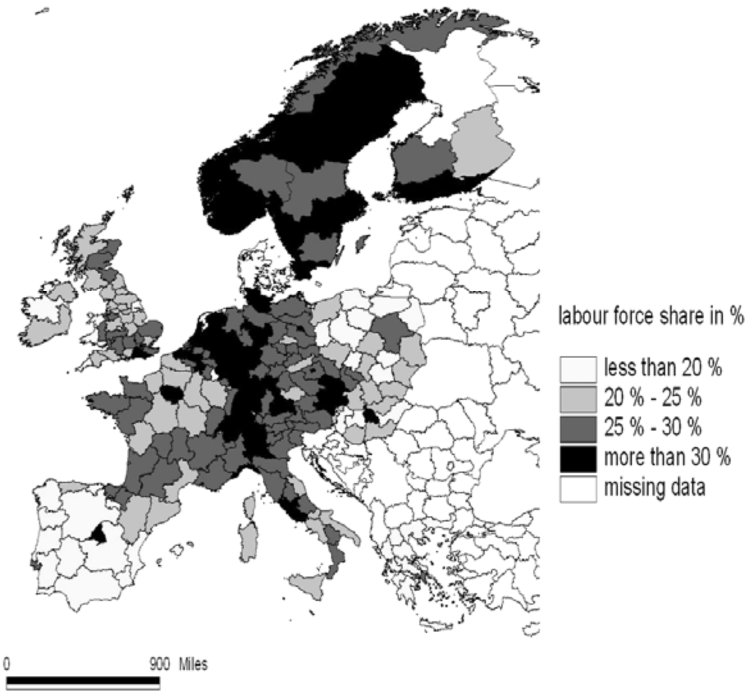
Sianesi, B., Van Reenen, J. (2003): The returns to education: Macroeconomics, Journal of Economic Surveys 17, 157-200.

Temple, J. (2001): Growth effects of education and social capital in the OECD countries, OECD Economic Studies 33, 57-101.

Figure 1: Univariate indicators for human capital

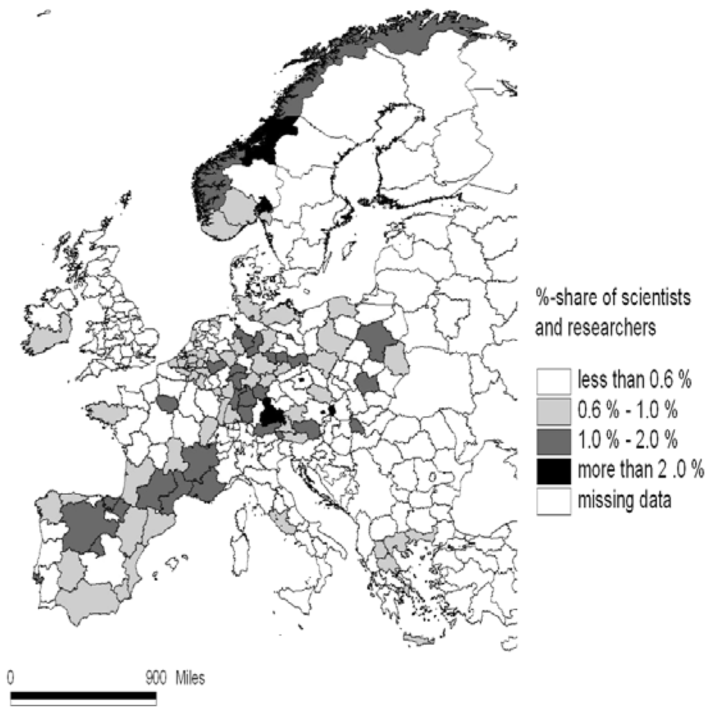
A People occupied in science and technology

-Percentage of regional labour force, 2005



B Scientists and researchers

-Percentage of regional employment, 2003



Source: Eurostat. Regional science and technology database.

Figure 2: Multivariate human capial indicator

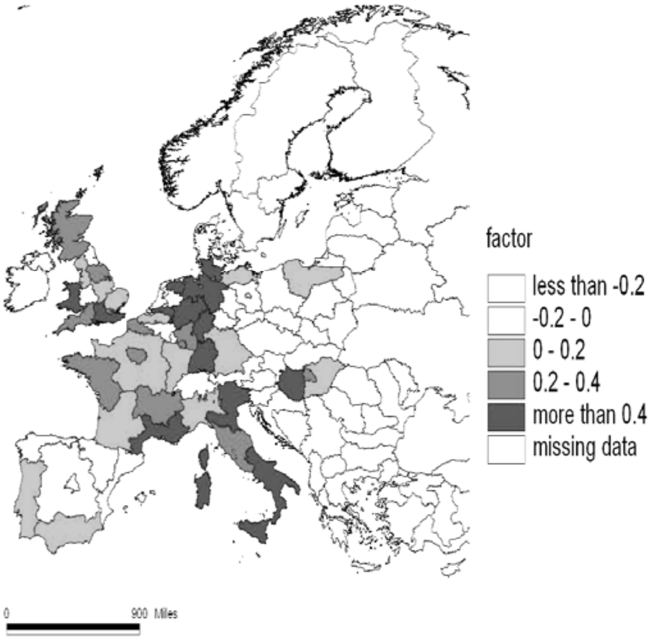


Table 1: Impact of researchers and scientists on economic growth

	High skilled workers	Income growth
Constant	21.649 (30.97)	0.019 (3.13)
Scientists and researchers	6.323 (8.58)	
High skilled workers (fitted)		0.125 (5.24)
Initial income per capita		-0.013 (11.18)
Adjusted R-squared	0.283	0.406

Note: Two step estimation. Human capital indicators from Eurostat, regional science and technology database. GDP per capita from Cambridge Econometrics. Initial income is gross value added in 1995. Income growth is mean annual growth rate over the 1995-2005 period. Results based on 185 NUTS2 regions, including 7 Norwegian regions.

Table 2: Ingredients of composite regional human capital indicators

Schooling: All, male, female, students at ISCED 3, ISCED 3 (GPV), ISCED 5-6 level

Human resources in science and technology: All, male, female, different age groups: 25-34, 35-44, 45-64, below 25 and over 64.

R&D expenditures: Private and government expenditures, universities.

Time period: Schooling variables 1998/99-2005, human resources in science in technology: 1996-2006, R&D expenditures: 1997-2003.

Regional availability of indicators: *Belgium*: Brussels, Vlaams Gewest, Wallonne. *Germany*: Baden-Wuerttemberg, Bayern, Berlin, Bremen, Hamburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Sachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen. *Spain*: Noroeste, Noreste, Comunidad de Madrid, Centro, Este, Sur, Canaris *France*: Île de France, Bassin Parisien, Nord–Pas-de-Calais, Est, Ouest, Sud-Ouest, Centre-Est, Méditerranée. *Italy*: Nord Ovest, Nord Est, Centro, Sud, Isole. *Hungary*: Közép-Magyarország, Dunántúl, Alföld és Èszak. *Netherlands*: Nord-, Zuid-Nederland. *Poland*: Centralny, Poludniowy, Wschodni, Pólnocno-Zachodni, Poloudniowo-Zachodni, Pólnocny. *Portugal*: Continente. *Finland*: Manner-Suomi. *United Kingdom:* North West, Yorkshire and The Humber, East Midlands, West Midlands, Eastern, London, South East, South West, Wales, Scotland, Northern Ireland.

Note: ISCED = International Standard Classification of Education. Level 3 refers to secondary education, and 5-6 to tertiary education.

Table 3: Reliability ratios for popular human capital measures

Regressor	Reliability ratio	t-value
Schooling (All students)	0.176	6.76
Human ressources in science and technology	0.191	6.54
R&D expenditures	0.091	3.00

Note: Regression results based on 64 NUTS1 regions.

Table 4: Wage regression

Log hourly wage	Women		Men	
	Wage	**participation**	**wage**	**participation**
	b/se	b/se	b/se	b/se
Cumulated part-time employment since the age of 25 (years)	-0.021**	0.108**		
	(0.006)	(0.007)		
Cumulated part-time employment since the age of 25 squared (years)	0.001**	-0.003**		
	(0.000)	(0.000)		
Cumulated unemployment since the age of 25 (years)	-0.049**	-0.137**	-0.022	-0.350**
	(0.009)	(0.007)	(0.041)	(0.016)
Cumulated unemployment since the age of 25 squared (years)	0.002**	0.003**	0.004*	0.013**
	(0.000)	(0.000)	(0.002)	(0.001)
Intermediate education	0.196**	0.171**	0.154**	0.056
	(0.027)	(0.045)	(0.032)	(0.049)
High education	0.494**	0.224**	0.406**	0.206**
	(0.030)	(0.049)	(0.039)	(0.054)
Intermediate education & lives in East Germany	-0.058	-0.173**	-0.011	-0.039
	(0.055)	(0.052)	(0.053)	(0.054)
High education & lives in East Germany	-0.066	0.127*	-0.071	0.089
	(0.053)	(0.061)	(0.056)	(0.066)
Age	0.089**	0.205**	0.026	0.235**
	(0.012)	(0.010)	(0.027)	(0.009)
Age squared	-0.001**	-0.002**	-0.000	-0.003**
	(0.000)	(0.000)	(0.000)	(0.000)
German nationality	0.121**	0.018	0.017	-0.038
	(0.031)	(0.058)	(0.037)	(0.061)
Married		-0.170**	**	0.073
		(0.037)		(0.047)
Number of children		-0.096**	**	0.002
		(0.017)		(0.018)
Firm-size				
Base category: Less than 20 employees				
GE 20 LT 200	0.156**		0.121**	
	(0.019)		(0.021)	
GE 200 LT 2000	0.242**		0.200**	
	(0.022)		(0.023)	
GE 2000	0.348**		0.264	
	(0.022)		(0.023)	
Self employed-without employees	-0.286**		-0.212	
	(0.047)		(0.048)	

Tabelle 4 (cont’d)

Federal States				
Base category: Nordrhein-Westfalia				
Schleswig-Holstein	-0.118 **		0.027	
	(0.045)		(0.049)	
Hamburg	0.021		0.007	
	(0.062)		(0.066)	
Niedersachsen	-0.035		-0.015	
	(0.030)		(0.031)	
Bremen	-0.092		0.009	
	(0.086)		(0.097)	
Hessen	0.048		0.013 *	
	(0.032)		(0.033)	
Rheinland-Pfalz	-0.048		-0.050	
	(0.034)		(0.035)	
Baden-Wue	0.043		0.067 **	
	(0.027)		(0.027)	
Bayern	0.048 +		0.036 **	
	(0.025)		(0.026)	
Berlin	-0.140 **		-0.126 **	
	(0.047)		(0.047)	
Mecklenburg-vorp	-0.111		-0.297 **	
	(0.071)		(0.070)	
Brandenburg	-0.233 **		-0.307 **	
	(0.063)		(0.060)	
Sachsen Anhalt	-0.267 **		-0.329 **	
	(0.060)		(0.057)	
Thueringen	-0.237 **		-0.339 **	
	(0.062)		(0.057)	
Sachsen	-0.235 **		-0.362 **	
	(0.056)		(0.052)	
Intercept	0.039	-3.777 **	1.804 **	-3.958 **
	(0.297)	(0.181)	(0.613)	(0.176)
Mills				
Lambda	-0.121		-0.729 **	
	(0.099)		(0.201)	
N	8214.000		7615.000	

Note: Analysis based on German SOEP data. Significance levels: + p<0.1, * p<0.05, ** p<0.01.

Tabelle 5: Average monthly income in German NUTS1 regions

	Mean	N
Schleswig-Holstein	1033,05	305
Hamburg	1355,89	139
Niedersachsen	1126,75	876
Bremen	1253,56	69
Nordrhein-Westfalen	1266,30	2150
Hessen	1299,81	739
Rheinl.-Pfalz, Saarland	1118,94	621
Baden-Wuerttemberg	1314,74	1277
Bayern	1233,54	1527
Berlin	1078,57	374
Mecklenburg-Vorpommern	819,29	230
Brandenburg	1066,45	379
Sachsen-Anhalt	942,18	407
Thueringen	1000,05	392
Sachsen	961,64	732
Total	1188,87	10217

Average monthly Income in EUR by low qualified workers
no job qualification

	Mean	N
Schleswig-Holstein	728,64	42
Hamburg	526,41	14
Niedersachsen	788,91	146
Bremen	783,02	14
Nordrhein-Westfalen	773,77	329
Hessen	709,14	99
Rheinl.-Pfalz, Saarland	528,43	104
Baden-Wuerttemberg	714,28	227
Bayern	657,73	243
Berlin	608,50	51
Mecklenburg-Vorpommern	316,34	23
Brandenburg	389,65	33
Sachsen-Anhalt	365,90	40
Thueringen	471,46	38
Sachsen	446,29	64
Total	679,99	1467