Discussion Papers No. 531, February 2008
Statistics Norway, Research Department

Øivind A. Nilsen, Arvid Raknerud, Marina Rybalka and Terje Skjerpen

Skill Composition: Exploring a Wage-based Skill Measure

Abstract:
Most studies of heterogeneous labor inputs use classifications of high skilled and low skilled based on workers' educational attainment. In this study we explore a wage-based skill measure using information from a wage equation. Evidence from matched employer--employee data show that skill is attributable to many variables other than educational length, for instance experience and type of education. Applying our wage-based skill measure to a TFP growth analysis, we find that the TFP residual decreases, indicating that more of the change in value-added is picked up by our skill measure than when using a purely education-based measure of skill.

Keywords: Skill composition, wages, TFP.

JEL classification: J31, D24, C23

Address: Øivind A. Nilsen (Corresponding author): Norwegian School of Economics and Business Administration, Hellevn. 30, N-5045, Bergen, Norway. E-mail: oivind.nilsen@nhh.no

Arvid Raknerud, Statistics Norway, Research Department, E-mail: arvid.raknerud@ssb.no

Marina Rybalka, Statistics Norway, Research Department, E-mail: marina.rybalka@ssb.no

Terje Skjerpen, Statistics Norway, Research Department, E-mail: terje.skjerpen@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

The most common way of accounting for labor heterogeneity is to classify workers as high skilled or low skilled based on their years of schooling. Another method is to assume that the relative efficiency of any two workers equals their wage ratio (see Griliches, 1960). Based on this assumption one may calculate efficiency-adjusted man-hours. Both methods have obvious shortcomings. Years of schooling may be too approximate a proxy for skill. Other observed and unobserved variables should also be taken into account (see the discussion in Borghans et al., 2001). Observed wage differences reflect not only skill differences, but also variables unrelated to skill, such as regional and temporal variations in labor market conditions, rent sharing, bargaining power, and transient fluctuations.

The current study looks at two different definitions of low skilled and high skilled. The first is the conventional definition based on years of education. In the other we utilize a wage equation framework and decompose a worker's wage into two parts: the first is a function of variables related to the worker's skill (observed and unobserved personal characteristics), while the second covers inter alia labor market and time specific characteristics and transient errors.[1] Each observation (i.e., a worker in a specific year) is then allocated to a skill group according to the size of the first part of the wage equation. We explore the implications of this wage-based skill measure for labor composition and relative wages in the firms. Furthermore, we adjust man-hours according to the worker's efficiency. To illustrate the importance of the choice of skill measure, we apply the different measures to analyze TFP growth. In accordance with our a priori belief, we obtain lower TFP growth when skill is represented by a wage-based skill measure rather than the length of education.

The paper proceeds as follows. In Section 2 we describe the data used in our analysis. Section 3 deals with classification of labor with respect to skill. In Section 4, we explore the implication of our wage-based skill measures for the growth in total factor productivity (TFP). Section 5 concludes the paper.

[1] Our method of formulating the wage equation has some similarity with the specifications used by Abowd et al. (1999), Iranzo et al. (2006) and Hellerstein and Neumark (2007), but is somewhat simpler since we do not explicitly account for firm effects. These are instead implicit in the transient noise term.

2 Data

To classify workers in a firm as high skilled or low skilled, we use matched employer–employee panel data for narrowly defined Norwegian manufacturing industries, covering the period 1995–2005. The two main data sources are the Register of Employer and Employees (REE) and the National Education Database (NED). The REE provide us with information on man-hours, wages — constructed as earnings divided by contracted annual working hours — and worker's place of residence; NED provides information on length and type of education. Worker's experience is calculated in the usual way as potential experience, i.e., as a person's age minus the length of education minus the age at which he/she started at compulsory primary school. When investigating TFP growth, we also utilize annual firm-level information from the accounting statistics on value-added and capital (for details about the capital variable see Raknerud et al., 2007) at the end of the year in constant prices. The sample is based on information from joint stock companies, which account for about 90 percent of total man-hours in manufacturing.[2]

3 Skill classification and construction of variables

We consider two ways of classifying a worker in a particular year as either low skilled or high skilled. According to the first definition, a worker is classified as high skilled if the length of his/her education is at least 13 years. According to the second definition, based on information from a wage equation, a worker is classified as high skilled in a period if the part of the predicted wage attributed to his/her personal characteristics equals or exceeds a certain threshold value. The threshold value equals the predicted wage (in that industry) of a hypothetical reference person with 13 years of education (see below). Thus, this definition utililises information in several variables and is specified as follows. For each industry consider the following wage equation:

$$\ln(W_{pt}) = X_{1pt}\gamma_1 + X_{2pt}\gamma_2 + \nu_p + \varepsilon_{pt}, \tag{1}$$

[2] For a more detailed description of data sources used, see the Data Appendix of Nilsen et al. (2006).

where W_{pt} is the hourly wage of person p in year t in a given industry. On the right hand side, we specify two (row) vectors with observed variables, X_{1pt} and X_{2pt}. X_{1pt} contains values of variables describing the individual's skill, i.e., the length of his/her education, experience, powers of experience up to the fourth order, type of education (represented by dummies) and gender. The vector has a time index, since some of its elements are allowed to change over time. X_{2pt} consists of year-specific dummies and dummies related to local labor market areas, i.e., observed variables that are assumed to be unrelated to an individual's skill.[3] The corresponding vectors of regression coefficients are denoted γ_1 and γ_2, respectively. The scalar ν_p is an unobserved random effect of individual p and ε_{pt} denotes a genuine error term. The unknown parameters in (1) are estimated by GLS using unbalanced panel data for each industry separately.

Since wages of part time workers are to some extent hampered by measurement errors, we only utilise data for full-time workers. In addition, we trim the data by utilizing quantile regressions for the 5 and 95 percent quantiles in each industry, with dummies for labor market regions and years as regressors. When estimating the wage equation (1), we omit observations that are characterized by either hourly wages below the conditional 5 percent or above the conditional 95 percent quantiles.[4]

The predicted log wage, $\widehat{\ln(W_{pt})}$, is decomposed into two parts. The first, denoted ω_{pt}, is relevant to skill measurement, while the second is related to the variables in the vector X_{2pt}, which are irrelevant to skill measurement. That is

$$\widehat{\ln(W_{pt})} = \omega_{pt} + X_{2pt}\widehat{\gamma}_2, \tag{2}$$

where

$$\omega_{pt} = X_{1pt}\widehat{\gamma}_1 + \widehat{\nu}_p.$$

A 'hat' above a parameter denotes an estimate, while $\widehat{\nu}_p$ is the predicted random effect based on the GLS estimation. We compare ω_{pt} with a threshold value, ω^{ref}, related to a hypothetical reference person, who we define as having 13 years of education and

[3] The definition of the seven labor market region dummies are based on characteristics such as size and centrality (see http://www.ssb.no/english/subjects/06/sos110_en/sos.110_en.pdf).

[4] The results from the wage equation estimations (available on request) show returns to education of approximately 5 percent, in line with other studies based on Norwegian data (see for instance Hægeland et al., 1999). The experience variable has a turning point around 30–32 years of experience, and the effect of gender, labor market dummies, and type-of-education dummies are all in line with our a priori expectations.

industry specific mean values (conditional on 13 years of education) for experience, type of education, gender and $\widehat{\nu}_p$. Since we correct for the effect of time and local labor market areas through X_{2pt}, the threshold value ω^{ref} has no subscripts. Our rule is that person p in period t is classified as high skilled if $\omega_{pt} \geq \omega^{ref}$, and low skilled in the opposite case.

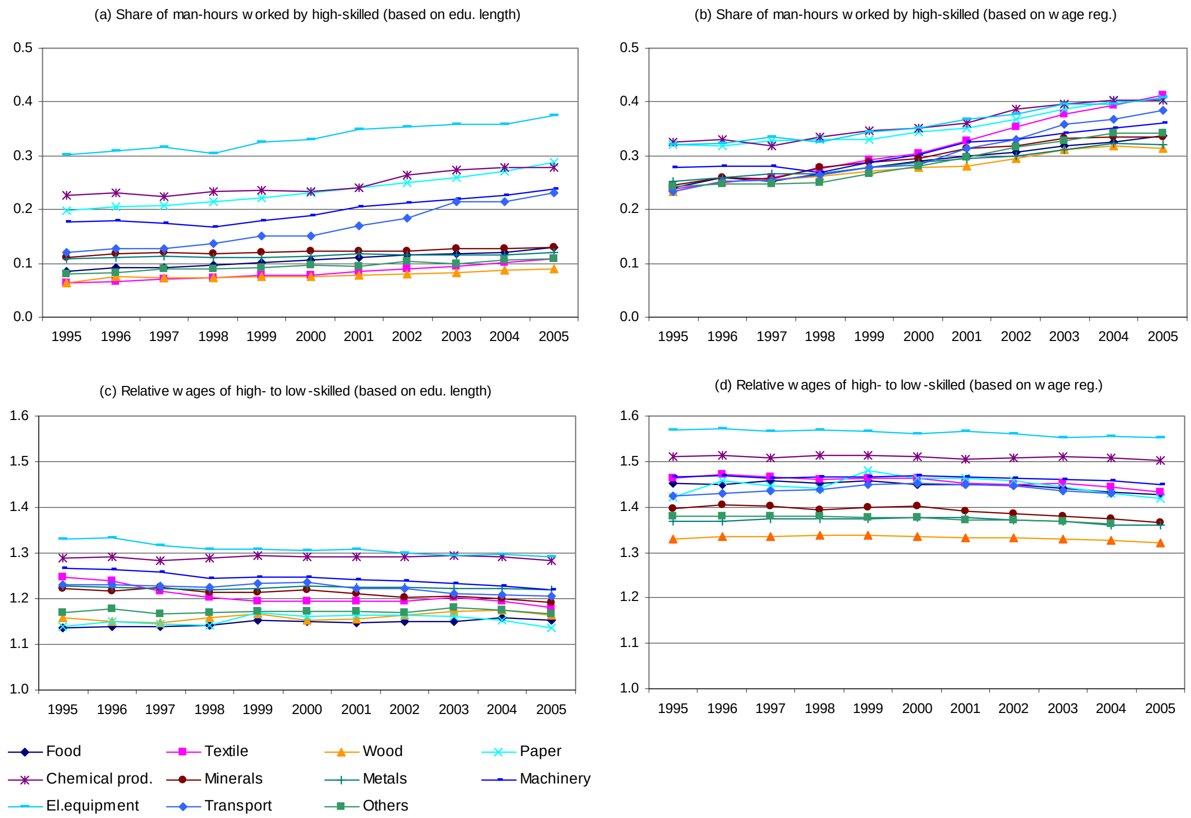

Figure 1: Proportion of man-hours and relative wages of high skilled workers in different industries and for different skill measures, 1995–2005 (weighted means)

In Figure 1, we compare the averages (weighted by man-hours) across firms in the same industry with respect to skill composition and relative wages using the two skill measures. There is an upward trend in the use of high-skilled workers, but relative wages are more or less constant. We also see that with the education-based skill measure, the proportion of high-skilled workers is much smaller relative to the case with our wage-based skill measure in most industries. This is due to the fact that experience plays an important role when one applies the wage-based skill measure. More detailed analysis shows that approximately 70–75 percent of the workers are classified into the same group by the two skill measures. The remaining 25–30 percent of the workers are mainly low educated with long experience, who are classified as high-skilled workers according to our wage-based skill measure. It is also interesting to note that the similarity between the two measures is most pronounced for Electrical equipment; a high tech industry (and the industry where the estimated return to education is the highest). The relative wage differences between high and low skilled are much smaller using the education-based skill measure compared to the one based on a wage equation. This indicates that skill premiums are not only attributable to length of education, but also to other variables such as experience, consequently these effects should be taken into account when identifying skills.

As a refined classification, we also divide workers within each of the two skill groups (low and high skilled) into subcategories according to their efficiency, assuming that workers within the two groups are perfect substitutes when adjusted for efficiency differences. Let M_{it}^h and M_{it}^l denote the input of high-and low-skilled man- hours in firm i in period t, respectively. Furthermore, let $M_{(k)it}^h$ and $M_{(k)it}^l$ denote the number of man-hours worked in subcategory k, where $k = 1, 2, ..., 5$, for high- and low-skilled workers, respectively. The categories are sorted in ascending order with respect to efficiency such that the least efficient workers are in subcategory 1, and each subcategory contains the same proportion of total man-hours (i.e., 20 percent). It follows that we have

$$M_{it}^m = \sum_{k=1}^{5} M_{(k)it}^m,\ m = h, l.$$

The efficiency-adjusted aggregate man-hours for the two groups h and l can then be written as

$$\widetilde{M}_{it}^m = \sum_{k=1}^{5} \lambda_k^m M_{(k)it}^m,\ \ \lambda_1^m < \lambda_2^m < ... < \lambda_5^m,\ \ m = h, l, \tag{3}$$

where λ_k^m ($k = 1, 2, ..., 5$; $m = h, l$) are efficiency parameters, which are calibrated as follows. Consider all the values of ω_{pt} occurring in our sample. Let ω_{pt}^m denote the skill-related part of the predicted log wage of person p when he/she is in skill group m. Within each skill group, we collect the values for all persons in all periods and sort them in ascending order and divide them into five *categories* of equal size, i.e., quintiles. Let $\omega_{(k)}^h$ and $\omega_{(k)}^l$ denote the median predicted wage in the k'th quintile for high-skilled and low-skilled workers, respectively. We then calibrate the efficiency parameters as

$$\lambda_k^m = \frac{\exp(\omega_{(k)}^m)}{\exp(\omega_{(1)}^m)},\ k = 1, ..., 5,\ m = h, l. \tag{4}$$

The formula for λ_k^m can be derived from the assumption of perfect substitution within skill group m (i.e., h or l), so that relative wage equals relative productivity of any two workers from different categories within the same skill group. On the other hand, high- and low-skilled workers are not assumed to be perfect substitutes.

The calculated values of λ_k^m for all the manufacturing industries are displayed in Figure 2. If we consider λ_5^m, which represents the relative wage of the most and the least effective workers within skill group m ($m = h, l$), one sees that the wage gap is generally larger for low-skilled than for high-skilled workers.

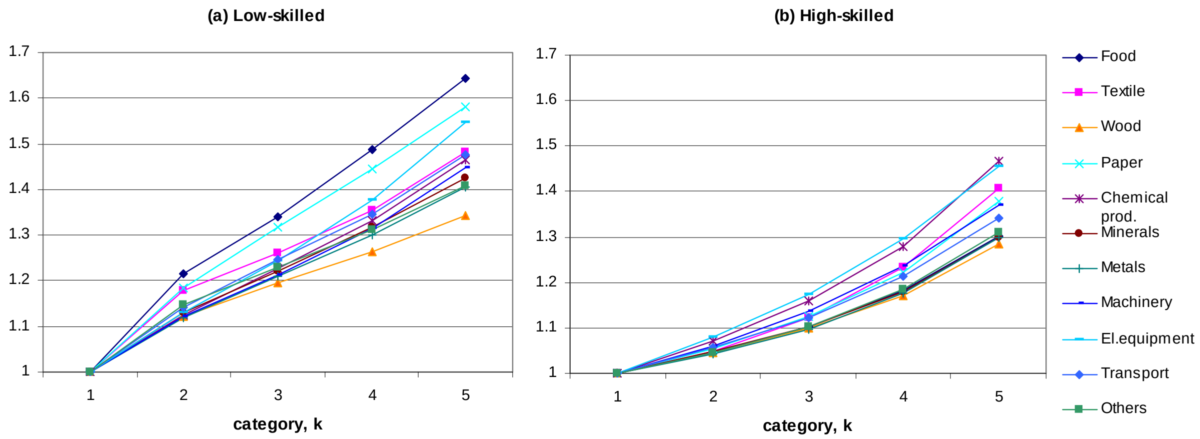

Figure 2: The efficiency parameters for low- and high-skilled workers in different industries

4 Productivity growth and different skill measures

Let us illustrate the importance of the skill measure by considering a simple example. Consider the following decomposition of the growth in labor productivity, $\Delta \ln(Y_t/L_t)$, at the industry level

$$\begin{aligned} \Delta \ln\left(\frac{Y_t}{L_t}\right) &= \alpha_{ht}\Delta \ln\left(\frac{\widetilde{M}_t^h}{L_t}\right) + \alpha_{lt}\Delta \ln\left(\frac{\widetilde{M}_t^l}{L_t}\right) \\ &\quad +(1-\alpha_{ht}-\alpha_{lt})\Delta \ln\left(\frac{K_t}{L_t}\right) + \Delta TFP_t, \end{aligned} \tag{5}$$

where Y_t, L_t and K_t denote value-added, total man-hours and the capital stock at the end of period t, respectively; and $\widetilde{M}_t^h$ and $\widetilde{M}_t^l$ are efficiency-adjusted man-hour aggregates at the industry level, defined as follows

$$\widetilde{M}_t^m = \sum_{i=1}^{N_t} \widetilde{M}_{it}^m,\ m = h, l, \tag{6}$$

where N_t denotes the number of firms in the industry in period t. Note that if $\lambda_k^m = 1$ for all k, $\widetilde{M}_t^m$ coincides with a non-efficiency-adjusted aggregate of man-hours in a given period, $M_t^m = \sum_{i=1}^{N_t} M_{it}^m$, $m = h, l$.

The term ΔTFP_t denotes growth in total factor productivity. For ease of exposition we suppress the index related to industry. We assume constant returns to scale, which means that the value of production equals total costs. The weights related to the two labor inputs are denoted by α_{ht} and α_{lt}, respectively. They are given as the arithmetic means of the income shares (the wage bill related to the skill group divided by value added in nominal terms) in the periods t and $t-1$.

For each industry in the manufacturing sector, we compare the TFP growth obtained from (5) with two other cases: First, when $\lambda_k^m = 1$ for all k and hence $\widetilde{M}_t^m, m = h, l$ in (5) are replaced by the non-efficiency-adjusted measures M_t^m, $m = h, l$. Second, when M_t^h and M_t^l are replaced by the corresponding education-based measure of high- and low-skilled man-hours. Note that the left-hand side of (5) does not depend on the skill measure used. Since we strongly believe that our wage-based skill measures represent labor input in a more appropriate way than the education-based one, our a priori belief is that the TFP growth will tend to be lower in the former case. This is because "explanatory power" is moved from the residual part to the first three factors in (5).

In Table 1 we report the mean annual growth in labor productivity over the period 1995–2005 together with the mean annual TFP growth according to three skill measures.[5] Corresponding to the last three columns in the table, we distinguish among three cases; the education-based skill measure, the wage-based skill measure without

[5] At the disaggregate manufacturing level there are arguments for using gross output instead of value-added as the output concept, as discussed in Jorgenson et al. (1987). Since we only consider an illustration, we retain value-added as the output concept at the disaggregate industry level. Furthermore, we do not consider the link between TFP growth at the plant/firm and the industry levels, as discussed in Hulten (2001, pp. 38–39).

Table 1: **Different skill measures and TFP growth**

Industry (NACE-codes)	Growth in labor productivity (%)	TFP growth (%) (1)	(2)	(3)
Food, beverages and tobacco (15-16)	2.79	1.06	0.95	0.83
Textile and leather products (17-19)	4.45	1.19	0.86	0.70
Wood and wood products (20)	4.00	2.35	2.24	2.18
Paper and publishing (21-22)	-0.00	0.41	0.34	0.28
Chemical and plastic products (23-25)	2.53	1.14	1.11	1.02
Mineral products (26)	0.86	0.36	0.28	0.24
Metal products (27-28)	5.58	1.97	1.91	1.87
Machinery (29)	3.57	0.53	0.48	0.39
Electrical equipment (30-33)	4.92	1.50	1.30	1.15
Transport and communication (34-35)	4.78	2.07	1.88	1.75
Furniture and others (36-37)	3.86	1.72	1.53	1.47
Average for manufacturing (15-37)*	3.22	1.26	1.16	1.07

Notes: All figures are simple means of annual growth rates in different productivity variables over 1995-2005. The TFP growth is calculated using eqs. (5) with different skill measures; the education length in (1), the wage-based skill measure in (2), and the wage-based skill measure with efficiency adjustment in (3). * Weights based on value added.

efficiency adjustment, and finally, the wage-based skill measure with efficiency differences within the two skill groups. Using the education-based skill measure, the mean annual TFP growth varies between 0.4 and 2.4 percent. In all the industries the mean annual TFP growth is lower using the two wage-based skill measures compared to the education-based skill measure. Looking at the TFP growth based on the two wage-based skill measures we find, which is reasonable, that the TFP growth is somewhat lower when we also adjust for efficiency differences within the two skill groups. Thus, the empirical results support our a priori beliefs that more appropriate ways of dealing with labor heterogeneity decrease TFP growth, since more of the change in value-added is picked up by the measurable components. The largest difference is found for Textile and leather products, where the TFP growth is 0.3 of a percentage point lower when using the wage-based skill measure without efficiency adjustment rather than the education-based skill measure. If one also accounts for efficiency differences within each of the two skill groups, TFP growth drops further by 0.2 of a percentage point. Also in Electrical equipment and Transport and communication we find a noticeable difference.

As a benchmark and in order to compare our results with what has been reported for other countries, we have calculated the TFP growth for the entire manufacturing industry, assuming homogeneous labor. We then obtain a TFP growth of 1.3 percent.[6] Compared with this benchmark, the TFP growth is 0.04 percentage points lower when heterogeneity in labor input is represented by length of education (see the last row of Table 1). The use of a wage-based skill measure with efficiency adjustment gives an additional decrease in the TFP growth of about 0.2 percentage points. An important question is then whether the mean TFP growth is statistically different when sampling uncertainty is taken into account. To answer this question we provide standard errors of the mean difference in TFP growth by means of bootstrapping.[7] We find that the difference in estimated TFP growth between our efficiency-adjusted wage-based skill measure and the education-based measure is statistically significant (the estimated standard error of the difference equals 0.08 percentage points). If we now consider a 50-years horizon as an example, which is quite common in long-run projections, a constant annual TFP growth rate of 1.07 instead of 1.26 percent implies a 10 percent lower TFP level after such a time span. Thus, an improved skill measure may have non-negligible effects especially in TFP accounting, where the effects accumulate over years.

5 Concluding remarks

In this paper we have constructed and elaborated on a wage-based skill measure, which is based on extracting information from a wage equation. We find that the relative wage differences between high-skilled and low-skilled workers are much smaller using only educational attainment for classification instead of our wage-based measure. This

[6] This growth rate is rather close to calculations using Norwegian national accounts data, which show an annual TFP growth for the manufacturing industry of 1.5 percent for the same period as we have studied. The EU-KLEMS project (see http://www.euklems.net/) reports (implicitly) that the (valued-added based) average TFP growth over the same period for a subgroup of the EU countries is on the short side of 1.

[7] The bootstrap works as follows. From the dataset used to produce the TFP growth estimates reported in Table 1 we draw a sample of N firms (with replacement). For each of these N firms we use the entire time series of output, wage costs, hours of work, and capital. In each replication we calculate the difference between the mean TFP growth using the wage-based skill measure with efficiency adjustment and the skill measure based on the length of education. After 250 bootstrap replications, we calculate the standard deviations of the differences in mean TFP growth over the bootstrap sample and take this as an estimate of the standard error of the difference in mean TFP growth.

indicates that skill is attributable to many other factors than educational length. We have applied our wage-based skill measure to TFP growth analysis. It appears that a wage-based skill measure that also accounts for efficiency differences within the two skill groups, is a more appropriate measure of skills than a traditional measure based on only educational attainment. In future research we will investigate whether the improved skill measure also aids our understanding of the increasing wage difference between high-skilled and low-skilled workers observed in many countries, and also its implication for heterogeneous labor demand at the firm level.

References

[1] Abowd, J.M., Kramarz, F. and D.N. Margolis, 1999, High Wage Workers and High Wage Firms, Econometrica 67, 251–333.

[2] Borghans, L., F. Green and K. Mayhew, 2001, Skill Measurement and Economic Analysis: An Introduction, Oxford Economic Papers 53, 375–384.

[3] Griliches, Z., 1960, Measuring Inputs in Agriculture: A Critical Survey. Journal of Farm Economics 42, 1411–1433.

[4] Hellerstein, J. and D. Neumark, 2007, Production Function and Wage Equation Estimation with Heterogeneous Labor: Evidence from a New Matched Employer-Employee Data Set, in: E.R. Berndt and C.R. Hulten, eds., Hard-to-Measure Goods and Services: Essays in Honor of Zwi Griliches, (Chicago University Press) 31–72.

[5] Hulten, C., 2001, Total Factor Productivity: A Short Biography, in: C.R. Hulten, E.R. Dean and M.J. Harper, eds., New Developments in Productivity Analysis, (Chicago University Press, Chicago) 1–47.

[6] Hægeland, T.J., T.J. Klette and K.G. Salvanes, 1999, Declining Returns to Education in Norway? Comparing Estimates across Cohorts, Sectors and Over Time, Scandinavian Journal of Economics 101, 555–576.

[7] Iranzo, S., F. Schivardi and E. Tosetti, 2006, Skill Dispersion and Firm Productivity: An Analysis with Employer–Employee Matched Data, CEPR Discussion Paper, no. 5539, London.

[8] Jorgenson, D.W., F.M. Gollop and B.M. Fraumeni, 1987, Productivity and U.S. Economic Growth, (Harvard University Press, Cambridge, MA).

[9] Nilsen, Ø.A., A. Raknerud, M. Rybalka and T. Skjerpen, 2006, Lumpy Investments, Factor Adjustments and Productivity, Discussion Paper No. 9/2006, Norwegian School of Economics and Business Administration, Bergen, Norway.

[10] Raknerud, A., D. Rønningen and T. Skjerpen , 2007, A Method for Improved Capital Measurement by Combining Accounts and Firm Investment Data, Review of Income and Wealth, 53, 397–421.