How Much Wage Variation Can We Explain?

Andreas Tiltag[*]

Department of Economics, University of Trier, Universitätsring 15, 54296 Trier, Germany

Abstract

Earnings functions are an important tool in labor economics as they allow to test a variety of labor market theories. Most empirical earnings functions research focuses on testing hypotheses about sign and magnitude for the variables of interest. In contrast, there is little attention for the explanation power of the econometric models employed. Measures for explanation power are of interest, however, for assessing how successful econometric models are in explaining the real world. Are researchers able to draw a complete picture of the determination of earnings or is there room for further theories leading to alternate econometric models? This article seeks to answer the question with a large microeconometric data set from Germany. Using linear regression estimated by OLS and R^2 as well as adjusted R^2 as measures for explanation power, the results show that up to 60 percent of wage variation can be explained using only observable variables.

Keywords: Labor economics, wage, theories of wage determination, earnings functions, wage variation, R^2, human capital, working conditions

JEL codes: J30, J31, J39

1. Introduction

Earnings functions are an important tool in labor economics. They allow to test a variety of labor market theories, to estimate wage differentials for working conditions, to predict individual earnings or to analyze wealth and income distributions, amongst others.

Most empirical earnings functions research follows only one of several possible directions, however. Papers start with a theory – either verbally or mathematically – and derive hypotheses about sign and magnitude for the variables of interest. Then, econometric models deliver evidence for or against the hypotheses. In contrast, there is little attention for the explanation power of the econometric models employed. While the question of how much wage variation can be explained is a natural one, often posed by beginning researchers, introductory text books point out that one should not pay too much attention on measures for

[*] Corresponding author. E-Mail address: tiltag.science@gmx.de.

explanation power such as R^2 or adjusted R^2 (Wooldridge 2003). This statement is helpful when the research question deals with hypotheses testing, because specification of an unbiased econometric model is more important in this case. Only for specific purposes, introductory text books emphasize measures of explanation power, for example when different econometric models are tested against each other. In this case, measures such as adjusted R^2 or information criteria (*AIC*, *BIC*) are employed.

Measures for explanation power are of wider interest, however, for assessing how close econometric models approach their target to explain the real world. Are researchers able to draw a complete picture of the determination of earnings or is there room for further theories leading to alternate econometric models? A common belief says that microeconometric wage equations possess little explanation power. For example, Bowles, Gintis and Osborne (2001) write, that only little wage variation can be explained – more specifically, below 50 percent.

This article poses the question of how much wage variation can be explained using only observable variables. The reason for using only observable variables lies in the measures for explanation power and in comparison purposes with other research. While econometric methods dealing with only observable variables such as linear regression estimated by OLS deliver standard measures for explanation power such as R^2, methods handling also unobservable variables frequently transform the econometric equation and either provide no useful measure for explanation power (e.g. restricted control function estimator) or measures which cannot be compared with R^2 (e.g. fixed effects dummy variable estimator). By using linear regression estimated by OLS and a large microeconometric data set from Germany which contains many variables eligible for wage regression, the GSOEP, this article gives a clear answer to the amount of explained wage variation and contradicts the statement of Bowles, Gintis and Osborne (2001).

2. Theory and Methodology

Theories of wage determination guide the construction of an empirical earnings function, which is used to estimate the amount of explained wage variation. The goal of this article is not to decide between more or less conflicting theories of wage determination, but to scrutinize how well the theories do in explaining observed wage variation. The explanation power depends on the considered explanation attempts, which in turn guide the selection of variables for the econometric wage equation. Therefore, this article considers more than one theory of wage determination and groups the theories in a first step.

Starting point in developing an empirical earnings function is the widely known Mincer wage equation. Mincer (1958, 1974) tries to estimate the impact of human capital on wages. The general form of the equation is $wage = f(X_1) + u$, where f is a function, X_1 a vector of human capital variables – skills that contribute to the production of goods and u a disturbance term.

The functional form of the dependent variable determines basic statistical properties of the wage equation and guides interpretation of the results. Economists widely agree on logarithmizing the wage variable. The first reason is a more comfortable interpretation of coefficients. First, independent variables obtain a percentage interpretation. Second, measurement units can be ignored, whereby comparisons among different currency areas are facilitated. Although important in hypotheses testing about sign and magnitude of coefficients, this feature of the log-transformation is of little interest here, for more details see Tiltag (2014). The second reason for logarithmizing the wage variable is better statistical properties. The distribution of the logarithmized wage resembles more closely a normal distribution than the non-transformed wage (Tinbergen 1957; Wooldridge 2003). The econometric equation transforms to $\log wage = f(X_1) + u$.

The functional form of the various independent variables among each other may be additive or non-additive. A production function with additive inputs usually leads to an econometric equation with additive inputs, for more details see Tiltag (2014). This corresponds to the efficiency-units interpretation of human capital variables: Employers only care about total efficiency units and can substitute between the different employee qualities (Welch 1969; Thaler and Rosen 1976; Hartog 1980). Mincer (1974) himself decides, based on an eyeball-test, for the separable econometric equation. This decision has guided much of the literature and leads to a practical econometric equation of the form $\log wage = \beta_0 + \beta_1 X_1 + u$, where β_0 is a constant and β_1 a set of human capital coefficients. The goal of this article is to estimate R^2 and adjusted R^2 using widely accepted methodologies, which is why we follow the specification of Mincer. Concerning the functional form of each independent variable for itself, there are lots of different specifications in the literature. Basically, we decide for widely accepted functional forms having in mind our data set. We come back to this topic in the data section.

Wage is not exclusively influenced by human capital. Therefore, further theories of wage determination are introduced in the following paragraphs. The theory of compensating wage differentials states, that wages are influenced by job variables. Examples are heavy lifting or fixed-term labor contracts. Prominent contributions come from Adam Smith (1776) and

Sherwin Rosen (1986). According to the basic principle, firms want some workers to execute jobs with unfavorable working conditions. As workers prefer jobs with favorable working conditions anything else equal, they demand a wage premium for unfavorable working conditions. Again, the widely used functional form is the separable one; per assumption, worker skills are not influenced by job characteristics (Thaler and Rosen 1976). The econometric equation extends to $\log wage = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, where β_2 is a set of job coefficients and X_2 a vector capturing the job variables.

Many articles skip a discussion for including employer variables. In fact, there are two lines of argumentation for their consideration. The first and more uncomplicated is to interpret them as working conditions. In this case, they can be treated analogously to the job variables. The second line relies on market imperfections, see for example Thurow (1975) or Krueger and Summers (1988). The econometric equation is now $\log wage = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$, where β_3 is a set of employer coefficients and X_3 a vector of employer variables.

The fourth set of standard variables in empirical wage equations is variables of the job environment. Basic parameters of the wage determination process may differ over region or time (Willis 1986). A subitem is to account for trends such as inflation in the wage equation, discussed by Altonji and Williams (2005). Again, the separable functional form is found most frequently, except if focus lies on examining the changing influence of the variables of interest above region or time, which is not of interest for this article. The resulting empirical wage equation is

$$(1) \qquad \log wage = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u.$$

The work horse methodology for estimating the equation is OLS. We follow this approach, which allows comparing results with a large number of other studies. As this article does not focus on unbiased coefficients, but on explained wage variation, we refrain from using estimators developed to account for unobserved variables such as fixed effects or instrumental variables. Neither of them delivers well-interpretable measures of explained wage variation.

The goal of the article is to estimate the amount of wage variation that can be explained with observed variables. To this end, several determination measures wait in the wings. The most common one is R^2 giving the fraction of variation in the dependent variable, which can be contributed to the variation in the independent variables (Wooldridge 2003). The percentage fraction results after multiplying with 100 (Wooldridge 2003). For reasons of completeness, we also state adjusted R^2 in the results tables. This measure is founded in the statistical properties of R^2: It gets the larger, the more variables enter the linear wage regression. Researchers prefer simpler models with fewer variables, anything else equal. Therefore, there is a tradeoff between explanation power and model complexity. Adjusted R^2 seeks to balance the two conflicting goals and penalizes every additional variable (Wooldridge 2003).

3. Data

3.1. Population and Sample

To gain a profound understanding of how much wage variation can be explained using only observable variables, we dispose two requirements on the data set. First, the data should cover a whole labor market in order to include the many facets of wage determination. Second, the data set should contain as many variables as possible capable of capturing the four sets of variables discussed in the theoretical section. The German Socio-economic Panel (GSOEP) fulfills both requirements. The GSOEP contains micro data for the German resident population (Haisken-Denew and Frick 2005). We use the waves from 1995 to 2007 and construct a pooled data set, consisting of the pooled cross-sections from 1995 to 2007. We thank the DIW Berlin for providing the data.

The research question suggests a definition of the population based on the labor force status, as the research question focuses on earnings which workers – as member of the labor force – receive. We use the definition of the International Labor Organization (ILO), further refined by the German Statistical Office (Statistisches Bundesamt 2007). Starting with the whole labor force, we include as many persons as possible in order to minimize problems of sample selection. However, we have to exclude some persons, because the neoclassical paradigm guiding the theoretical considerations does not apply for them or because they do not provide information for all variables used in this article. From the labor force, we exclude the unemployed as well as the self-employed, as they do not receive wages, so no wage variation can be explained. Besides, only two groups of persons have to be excluded: Special groups (persons in maternity leave, partial retirement with zero working time, military service, civil service and voluntary ecological year) and apprentices, both because of missing variables. Removing additionally all persons with missing values on any of the variables employed, we remain with a sample of 14,093 workers.

3.2. *Dependent Variable*

Ideally, the wage variable contains all components of the bundle (in currency units) workers receive from their employers: Employers use the labor power of workers to produce goods and offer workers wages in exchange. We have to exclude special payments such as a thirteenth or fourteenth monthly payment in a year, Christmas bonus or holiday pay, as they cannot be matched with a specific job; the GSOEP provides information only about special payments for the whole past year – in which a worker can hold either one or more than one job. Note, that our approach is in line with the literature because in most data sets, special payments cannot be uniquely matched to one job (Thaler and Rosen 1976; Altonji and Williams 2005). Researchers usually use gross wages, which we follow. Employers can produce with the same labor units more goods, if they can use the resources for a longer period. Therefore, we use wages per hour as dependent variable. To derive the dependent variable, we divide wages per month by the product of the average realized working hours per week and the average weeks per month (4.3452). Finally, real wages instead of nominal wages are mostly employed in the literature to account for inflation. We deflate gross hourly wages by the consumer price index of the German Statistical Office with the base year 2005 (Statistisches Bundesamt 2011).

3.3. *Independent Variables*

The classical human capital variables used by Mincer are education, which reflects general human capital from formal training, and labor market experience as well as labor market experience squared, which reflect general human capital from on-the-job training. Going beyond Mincer and accounting for skill-based human capital, we define a set of occupation dummies based on the Klassifikation der Berufe 1998 (KldB 1998). Another group of variables potentially reflecting human capital are demographics. First, firms can utilize human capital only if it is available (Becker 1962; Grossman 1972). Workers with health impediments could suffer from underutilization of their full human capital potential. Therefore, we include a set of dummy variables for health status and one dummy variable for disability. Second, married workers may be more productive (Hellerstein, Neumark and Troske 1999), either because of a causal effect or because of selection into marriage. We include the two dummy variables married and was married with the comparison group not married. Third, females and foreigners may differ in their productivity or may be discriminated against – a discussion which we do not pursue here. However, as in many wage regressions we consider a female dummy and a foreigner dummy.

Human capital encompasses a vast amount of manifestations. While the variables specified above are more or less standard in empirical wage equations, researchers stress the importance of unobserved human capital variables. Neglecting them can lead to biased results. The most prominent unobserved human capital variables are those for unobserved abilities. As the focus of this article lies not on coefficients but on explained wage variation, unobserved ability is not as important as in other articles. Nevertheless, we try to account for otherwise unobserved ability by using two groups of variables for two reasons: First, readers are more comfortable with unbiased results. Second, these variables may deliver additional explanation power.

The first group of variables for unobserved ability shall account for intelligence. In the 2006 wave of the GSOEP, intelligence is measured with two ultra-short intelligence tests derived from life-span psychology (Baltes, Lindenberger and Staudinger 1998; Lindenberger 2002). According to the theory, all cognitive abilities can be grouped into mechanic (fluid) and pragmatic (crystallized) intelligence. As a proxy for fluid intelligence, perceptual speed was chosen. The Digit Symbol Modality Test from the Wechsler Adult Intelligence Scale (WAIS) was modified and the Digit Symbol Test (DST) developed (Lang, Weiss, Stocker & von Rosenbladt 2007).[1] As a proxy for crystallized intelligence, word fluency was selected. The Animal-Naming-Task (ANT) was adapted (Lang, Weiss, Stocker and von Rosenbladt 2007).[2] Overall, both tests reach an acceptable reliability and a sufficient validity, whereby the DST performs better than the ANT (Lang, 2005; Lang, Weiss, Stocker and von Rosenbladt 2007). The intelligence variables are only surveyed in 2006, so we have to assume their stability over time and expand the 2006 values to the other years.

The second group of variables for unobserved ability measures personality. This article relies on the Big 5 concept. While originally, a long questionnaire is used to measure the Big 5, the 2005 wave of the GSOEP relies on a short version of 15 questions for reasons of tractability. Respondents state their answers on a 7-point Likert-scale ranging from 1 “I strongly disagree” to 7 “I strongly agree”. Reliability and validity for the short questionnaire

---

[1] Survey participants see a symbol on a computer screen and have to type the corresponding digit, whereby the translation from symbol to digit remains visible all the time. After 90 seconds, the test automatically ends and the number of correct answers is calculated.

[2] Survey participants have to name as many different animals as possible. Again, the test stops after 90 seconds and the number of correct answers is calculated.

for the Big 5 reach satisfactory results (Gerlitz and Schupp 2005). Factor analysis is used to extract the five personality traits openness, extraversion, agreeableness, conscientiousness and neuroticism. See Dehne and Schupp (2007) for further details. Because the Big 5 are only measured in 2005, we have to assume their stability over time, again. Wichert and Pohlmeier (2010) find some evidence supporting the assumption. The Big 5 variables of 2005 are passed on the other waves.

Among the job variables is one more group of human capital variables: Tenure and tenure squared are considered to measure specific human capital from on-the-job training (Mincer and Jovanovic 1981). Coming to the other job variables, required training captures eventually higher demanded effort levels in the job. The white collar-variable goes in the same direction, which reflects an eventually lower physically effort and a higher mental effort. Dummies for part-time employment, marginal employment, a fixed-term labor contract and job-creating measures capture the flexibility and risk connected with the specific job arrangement.

In total, we include three groups of employer variables. They encompass employer size (Brown and Medoff 1989). To obtain the same groupings in every year of the GSOEP, we specify four dummy variables: The first reaches from employer sizes 1 to 19 employees, the second from 20 to 199, the third from 200 to 1,999 and the fourth starts with 2,000 employees. Industry is the other standard variable (Krueger and Summers 1988; Gibbons and Katz 1992). Based on the Klassifikation der Wirtschaftszweige 2003 (WZ 2003), which derives in turn from the NACE Rev. 1.1, we specify ten industry dummies summarized in Table 1. Besides firm size and industry, we use the variable public, stating whether the firm is in the public sector or not.

Table 1. Summary Statistics

Variable	Mean	SD	Minimum	Maximum
Wage				
log(gross wage per hour)	2.495	0.479	-1.148	5.925
Gross wage per hour	13.523	7.610	0.317	374.316
Employee				
Education	11.847	2.442	7.000	18.000
Experience	18.749	10.946	0.000	53.500
Experience^2	471.310	469.288	0.000	2862.250
Occupation				
Construction, materials	0.062	0.241	0.000	1.000
Manufacturing	0.113	0.317	0.000	1.000
Electric	0.019	0.137	0.000	1.000
Technology	0.065	0.247	0.000	1.000
Trade, art	0.144	0.351	0.000	1.000

Variable	Mean	SD	Minimum	Maximum
Basic services	0.155	0.362	0.000	1.000
Business, administration	0.196	0.397	0.000	1.000
Order, security	0.044	0.205	0.000	1.000
Health	0.064	0.244	0.000	1.000
Education	0.079	0.270	0.000	1.000
Other	0.020	0.140	0.000	1.000
Health status				
Less good	0.094	0.292	0.000	1.000
Satisfactory	0.303	0.460	0.000	1.000
Good	0.480	0.500	0.000	1.000
Very good	0.109	0.312	0.000	1.000
Disabled	0.069	0.254	0.000	1.000
Female	0.471	0.499	0.000	1.000
Marital status				
Married	0.668	0.471	0.000	1.000
Was married	0.120	0.325	0.000	1.000
Foreigner	0.111	0.314	0.000	1.000
Verbal fluency	103.779	14.646	80.918	168.465
Perceptual speed	104.029	13.990	80.379	138.828
Openness	51.320	9.540	10.685	74.576
Extraversion	49.131	10.080	19.465	85.326
Conscientiousness	49.628	9.571	16.927	81.270
Agreeableness	50.597	9.631	18.686	88.001
Neuroticism	48.445	9.392	6.220	91.362
Job				
Required training				
Vocational training	0.567	0.496	0.000	1.000
Higher education	0.135	0.342	0.000	1.000
Tenure	10.861	9.959	0.000	48.100
Tenure^2	217.141	338.828	0.000	2313.610
White-collar	0.621	0.485	0.000	1.000
Fixed-term contract	0.072	0.259	0.000	1.000
Part-time	0.198	0.399	0.000	1.000
Marginal employment	0.053	0.224	0.000	1.000
Job-creating measures	0.007	0.084	0.000	1.000
Employer				
Employer size				
1–19 employees	0.247	0.431	0.000	1.000
20–199 employees	0.282	0.450	0.000	1.000
200–1,999 employees	0.225	0.418	0.000	1.000
Industry				
Primary	0.017	0.130	0.000	1.000
Manufacturing, utilities	0.257	0.437	0.000	1.000
Transportation	0.203	0.402	0.000	1.000

Variable	Mean	SD	Minimum	Maximum
Accommodation, food	0.016	0.125	0.000	1.000
Finance	0.042	0.200	0.000	1.000
Business	0.068	0.253	0.000	1.000
Administration	0.137	0.344	0.000	1.000
Education	0.061	0.240	0.000	1.000
Health	0.112	0.315	0.000	1.000
Other	0.024	0.152	0.000	1.000
Public	0.271	0.444	0.000	1.000

Notes: Sample encompasses all employed persons excluding the self-employed, apprentices, persons in maternity leave, partial retirement with zero working time, military service, civil service and voluntary ecological year. 14,093 observations.

Source: GSOEP (1995–2007), own calculations.

The job environment variables include country and time dummies. Country dummies for the German states (Bundesländer) account for the different economic conditions across regions, while time dummies account for inflation and economic trends, amongst others.

Summing up, we employ most of the standard variables used in empirical wage equations. Eventually, on-the-job-training is missing. Additionally, we can partly account for unobserved ability by including the intelligence and the Big 5 variables. Of course, there remain unobserved variables which influence wages. The determination of how far one gets with a thoroughly specified sample – remember the guideline to use as many worker observations as possible – and observable variables in explaining wage variation is exactly the goal of this article.

4. Results

All estimates of the amount of wage variation that can be explained use equation (1) in one form or another. A natural starting point is the Mincer equation, considering only the classical human capital variables education, experience and experience squared. As Table 2 shows, the approach can only explain 16.4 percent of the variation in the wage variable. Another obvious question to ask is, whether innate ability is a better predictor. For this reason, we only account for the IQ measures and the Big 5 variables. The results in column (2) of Table 2 reveal an even smaller magnitude of explained wage variation of a meager 4.4 percent.

Summing up these first results, a move beyond the original Mincer equation by considering additional variable groups is reasonable, because it seems that the Mincer approach captures only few influences involved in the wage determination process.

Table 2. Explained Wage Variation (OLS) by Gropus of Variables - Innate Ability and Original Mincer. Dependent Variable: log(Gross Wage per Hour)

Variable	(1) Original Mincer	(2) Innate Ability
Constant	1.380	2.195
	(0.023)	(0.052)
R^2	0.164	0.045
Adjusted R^2	0.164	0.045
Observations	14,093	14,093

Notes: Sample encompasses all employed persons excluding the self-employed, apprentices, persons in maternity leave, partial retirement with zero working time, military service, civil service and voluntary ecological year. Control variables include for innate ability: Verbal fluency, perceptual speed, openness, extraversion, conscientiousness, agreeableness and neuroticism; for original Mincer: Education, experience and experience^2.

Source: GSOEP (1995–2007), own calculations.

As mentioned in the theoretical section, several theories of wage determination coexist. To gain a better impression of the role of the different variable groups, Table 3 accounts for each group on its own. The amount of explained wage variation is lowest with 6.2 percent considering only market variables, shown in column (4). This is not astonishing, as their role is merely to capture different environmental influences. A noticeably larger fraction of 20.2 percent can be explained by employer variables in column (3). However, more important are human capital variables. In extension of the Mincer equation, researchers used other human capital variables as well as demographics in order to explain a larger fraction of wage variation. Moreover, we include innate ability via the IQ and Big 5 measures. Column (1) depicts an explanation power of 31.2 percent. Astonishingly, considering only the job variables in column (2) delivers the largest fraction of explained wage variation. These variables are able to explain a good 35.9 percent of wage variation.

Of course, the largest fraction of wage variation can be explained by considering all variable groups together. The results in column (5) of Table 3 show a remarkable 53.6 percent. This is an amount larger than vast parts of the literature deliver. We do not give examples from the literature here, because the large amount of studies estimating empirical wage equations makes any selection arbitrary. Our results outdate the statement of Bowles, Gintis and Osborne (2001), that only less than 50 percent of wage variation can be explained. Problems such as specifying the wage equation, the data or insufficient available observations may be reasons for a small amount of explained wage variation.

Table 3. Explained Wage Variation (OLS) by Gropus of Variables - Employee, Job, Employer and Market. Dependent Variable: log(Gross Wage per Hour)

Variable	(1) Employee	(2) Job	(3) Employer	(4) Market	(5) All Groups
Constant	1.596	2.178	2.757	2.438	1.900
	(0.066)	(0.010)	(0.015)	(0.031)	(0.062)
R^2	0.312	0.359	0.202	0.062	0.536
Adjusted R^2	0.311	0.359	0.202	0.060	0.533
Observations	14,093	14,093	14,093	14,093	14,093

Notes: Sample encompasses all employed persons excluding the self-employed, apprentices, persons in maternity leave, partial retirement with zero working time, military service, civil service and voluntary ecological year. Control variables include for employee: Education, experience, experience^2, verbal fluency, perceptual speed, openness, extraversion, conscientiousness, agreeableness and neuroticism and dummies for occupation, health status, disabled, female, marital status and foreigner; for job: Tenure, tenure^2 and dummies for required training, white-collar, public, job-creating measures, fixed-term contract, part-time and marginal employment; for employer: Dummies for employer size and industry; for market: Dummies for German states (Bundesländer) and year.

Source: GSOEP (1995–2007), own calculations.

5. Sensitivity Analysis

As there are many different specifications of the wage equation in the literature, their impact on the amount of explained wage variation is of interest. Table 4 in its first section provides information about different sample delineations. Frequent candidates for limiting the sample are excluding low wages, restricting the sample to workers between 18 and 65, excluding foreigners or workers in the primary sector and focusing on the private sector or West Germany. The general picture of sample restrictions is, that the amount of explained wage variation rises slightly, up to 55.1 percent selecting only the private sector. This seems plausible, as unexplained variation between non homogeneous groups may be reduced.

An even larger increase in explained wage variation results from giving up the pooled cross-sections and focusing on single years. In 1995 as well as in 2007, a share of about 60.0 percent of wage variation can be explained using only observable variables. Explanations of this result may lie in the consideration of the time variables in the pooled cross-sections. Dummy variables for years cannot account for changing impacts of the variables over the years. For example, the education coefficient may rise due to increasing human capital demands of employers. With the restriction on a single cross-section, the problem becomes obsolete.

The second section of Table 4 provides information about changing the variables in the wage equation. Giving up the advantages of the logarithmized wage, a drop to 35.4 percent in explained wage variation results. While transforming the dependent variable has a major impact, modifications in the list of only few independent variables cause little changes in the results. Dropping two variable groups, which are not widely available – the occupation dummies and the innate ability variables for IQ and Big 5 – the amount of explained wage variation drops only slightly, to 52.6 percent and 53.3 percent, respectively.

Table 4. Explained Wage Variation (OLS), Sensitivity Analysis. Restricted Sample, Cross Sections 1995 and 2007, Unlogarithmized Wage

	R^2	Adjusted R^2	Observations
	Sample		
Without low wages	0.543	0.541	13,551
From 18 to 65 years old	0.536	0.534	14,017
Without foreigners	0.544	0.541	12,534
Without primary sector	0.536	0.533	13,850
Private sector	0.551	0.548	10,280
West Germany	0.517	0.514	12,242
1995	0.600	0.546	566
2007	0.599	0.578	1 ,313
	Variables		
Unlogarithmized wage	0.354	0.350	14,093
Without occupation	0.526	0.524	14,093
Without innate ability	0.533	0.531	14,093

Notes: Sample encompasses all employed persons excluding the self-employed, apprentices, persons in maternity leave, partial retirement with zero working time, military service, civil service and voluntary ecological year. Control variables include for employee: Education, experience, experience^2, verbal fluency, perceptual speed, openness, extraversion, conscientiousness, agreeableness and neuroticism and dummies for occupation, health status, disabled, female, marital status and foreigner; for job: Tenure, tenure^2 and dummies for required training, white-collar, public, job-creating measures, fixed-term contract, part-time and marginal employment; for employer: Dummies for employer size and industry; for market: Dummies for German states (Bundesländer) and year.

Source: GSOEP (1995–2007), own calculations.

6. Conclusion

This article examined the question of how much wage variation can be explained using only observable variables. The question relates to the deeper question, whether well-known theories of wage determination are able to explain real world wage determination or whether the theories have to be developed further or even replaced. According to the results, the statement of Bowles, Gintis and Osborne (2001), that empirical wage equations can only explain below 50 percent of wage variation, is no longer valid. Relying on a proxy strategy for unobserved ability, the specifications in this article can explain more than 50 percent of the wage variation. A key is to account for all theories of wage determination and to find proxy variables for unobserved ability as well as using more measures for human capital than just the broad education and experience variables. Here, we used occupation which can account for task-based human capital. Moreover, the results reveal, that researchers can move the explained wage variation slightly upward when they restrict the sample with the effect of creating a more homogeneous group of workers. Finally, an increase to even 60.0 percent of explained wage variation results when only one cross-section is used.

In summary, this article delivers evidence, that available wage theories are at least able to explain a good part of observed wage variation. A more recent theory of wage variation, that delivers a more detailed view on the wage mechanisms acting in the labor market, is the Theory of Compensating Wage Differentials on Segmented Labor Markets (Tiltag 2014).

Acknowledgement

We thank the DIW Berlin for providing the data.

References

Altonji, Joseph G., & Williams, Nicolas (2005). Do Wages Rise with Job Seniority? A Reassessment. *Industrial and Labor Relations Review*, *58(3)*, 370–397.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (1998). Life-span Theory in Developmental Psychology. In R. M. Lerner (Ed.), *Handbook of Child Psychology*, 5th ed., Vol. 1 (pp. 1029–1143). New York: Wiley.

Becker, Gary S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, *70(5)*, 9–49.

Bowles, Samuel, Gintis, Herbert, & Osborne, Melissa (2001). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature*, *39(4)*, 1137–1176.

Brown, Charles, & Medoff, James (1989). The Employer Size-Wage Effect. *Journal of Political Economy*, *97(5)*, 1027–1059.

Dehne, Max, & Schupp, Jürgen (2007). Persönlichkeitsmerkmale im Sozio-oekonomischen Panel (SOEP) – Konzept, Umsetzung und empirische Eigenschaften. *Research Note 26*. DIW Berlin.

Gerlitz, Jean-Yves, & Schupp, Jürgen (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. *Research Note 4*. DIW Berlin.

Gibbons, Robert, & Katz, Lawrence F. (1992). Does Unmeasured Ability Explain Inter-Industry Wage Differentials? *Review of Economic Studies*, *59(3)*, 515–535.

Grossman, Michael (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, *80(2)*, 223–255.

Haisken-DeNew, John P., & Frick, Joachim R. (2005). *DTC. Desktop Companion to the German Socio-Economic Panel (SOEP)*. Berlin: Deutsches Institut für Wirtschaftsforschung.

Hartog, Joop (1980). Earnings and Capability Requirements. *Review of Economics and Statistics, 62(2)*, 230–240.

Hellerstein, Judith K., Neumark, David, & Troske, Kenneth R. (1999). Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations. *Journal of Labor Economics, 17(3)*, 409–446.

Krueger, Alan B., & Summers, Lawrence H. (1988). Efficiency Wages and the Inter-Industry Wage Structure. *Econometrica*, *56(2)*, 259–293.

Lang, F. R., Weiss, D., Stocker, A., & von Rosenbladt, B. (2007). Assessing Cognitive Capacities in Computer-assisted Survey Research: Two ultra-short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP). *Schmollers Jahrbuch*, *127*, 183–192.

Lindenberger, U. (2002). Erwachsenenalter und Alter. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie*, 5th ed. (pp. 350–391). Weinheim: Beltz PVU.

Mincer, Jacob (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.

Mincer, Jacob (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy, 66(4)*, 281–302.

Mincer, Jacob, & Jovanovic, Boyan (1981). Labor Mobility and Wages. In: Sherwin Rosen (Ed.), *Studies in Labor Markets* (pp. 21–64). Chicago, London: University of Chicago Press.

Rosen, Sherwin (1986): The Theory of Equalizing Differences. In: Orley Ashenfelter & Richard Layard (Eds.), *Handbook of Labor Economics, Vol. 1* (pp. 641–692). Amsterdam et al.: North-Holland.

Smith, Adam (1776): *An Inquiry into the Nature and Causes of the Wealth of Nations*. *Vol. 1*. London: Strahan, William, & Cadell, Thomas.

Statistisches Bundesamt (2011). *Preise 2010. Verbraucherpreisindizes für Deutschland. Jahresbericht (Januar 1991–Dezember 2010)*. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt (2007). *Statistisches Jahrbuch 2007. Für die Bundesrepublik Deutschland*. Wiesbaden: Statistisches Bundesamt.

Thaler, Richard H., & Rosen, Sherwin (1976). The Value of Saving a Life: Evidence from the Labor Market. In: Nestor E. Terleckyj (Ed.), *Household Production and Consumption* (pp. 265–302). New York: National Bureau of Economic Research.

Thurow, Lester C. (1975). *Generating Inequality. Mechanisms of Distribution in the U.S. Economy*. New York: Basic Books.

Tiltag, Andreas (2014). *Kompensierende Lohndifferentiale auf Segmentierten Arbeitsmärkten*. Göttingen: Optimus.

Tinbergen, Jan (1957). Welfare Economics and Income Distribution. *American Economic Review*, *47(2)*, 490–503.

Welch, Finis (1969). Linear Synthesis of Skill Distribution. *Journal of Human Resources*, *4(3)*, 311–327.

Wichert, Laura, & Pohlmeier, Winfried (2010). Female Labor Force Participation and the Big Five. *ZEW Discussion Papers, No. 10-003*.

Willis, Robert J. (1986). Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions. In: Orley Ashenfelter, & Richard Layard (Ed.), *Handbook of Labor Economics*, *Vol. 1* (pp. 525–602). Amsterdam et al.: North-Holland.

Wooldridge, Jeffrey M. (2003). *Introductory Econometrics*. *2. Ed.* Mason: Thomson South-Western.