International Journal of Economic Perspectives, 2017, Volume 11, Issue 2, 225-237.

Reexamining the Mincerian Wage Equation in Turkey: Functional form and Interpretation

Ebru C. AKAY
Department of Econometrics, Faculty of Economics, Marmara University, Goztepe Campus, Kadıkoy, Istanbul, Turkey. E-mail: ecaglayan@marmara.edu.tr.

Sinem G. KANGALLI UYAR*
Department of Econometrics, Faculty of Economics and Administrative Sciences, Pamukkale University, Kınıklı, Denizli, Turkey. E-mail: skangalli@pau.edu.tr. Tel: +90 541 855 7135, Fax: +90 258 296 2626.

ABSTRACT

The Mincer Wage equation is the mostly used human capital model. Since this model was introduced, there has been a substantial expansion in micro data and estimation techniques avaliable to labor economists. One of the aims of this study was that explaning how the Mincer equation stands in light of these advances in empirical labor economics. For this aim, the wage models were estimated in semiparametric form based on public and private data obtained from Household Labor Force Survey micro data set of 2013 in Turkey in order to analyze the empirical performance of mincerian wage model. The other aim of the study was that taking in consideration the spatial effect on wages. As differently from the other studies, coordinate variables were used in order to represent the spatial effect. The estimation results showed that there is a linear relationship between wage and experience for private sector in Turkey as different from the standard mincerian wage model. On the other hand, the coordinate variables measuring the spatial effect on wages showed that wages are maximum for public sector in Ankara and nearby cities. The wages are maximum across the country and for private sector in Istanbul and nearby cities with some provinces from the east of Turkey.

JEL Classification: C10; C14; J31.

Keywords: Semiparametric Regression; Mincerian Wage Equation; Penalized Iteratively Reweighted; Least Squares; Turkey.

**Corresponding author.*

1. INTRODUCTION

Assuming linearity when the relationship between X and Y is actually nonlinear can cause a misspecification problem. In this case, an analyst might conclude that there is no relationship between X and Y, when the two are strongly related. Keele (2008) stated that instead of assuming that we know the functional form for a regression model, a better alternative is to estimate the appropriate functional form from the data. In the absence of strong theory for the functional form, this is often the best way to proceed. To estimate the functional form from data, we must replace global estimates with local estimates. With global estimates, the analyst assumes a functional form for the model; with local estimates, the functional form is estimated from the data. The local estimators are referred to as nonparametric regression models or smoothers. What is needed is a multivariate model that allows one to combine global and local estimates. The solution is the semiparametric regression model. The aim of this study was to find out the best mincerian wage model based on sector fitting perfectly the data set obtained from Household Labour Survey in 2013 for Turkey. For this aim, semiparametric regression model determining the functional form through distribution of data set without any restrictive assumptions about functional forms of the relationships between variables was used. While the qualitative variables were added parametrically into the model, quantative variables added as nonparametrically in semiparametric regression model.

After the estimation of the semiparametric regression model, figures were obtained demonstrating functional forms between dependent variable and quantitative independent variables. In the study, the wage models were estimated seperately based on private and public sector employees beside the estimations obtained by using full data set.

Moreover, wage models were estimated in fully parametric form and semiparametric form. Then, it was examined that which model explains better the relationship by using LR tests. The other aim of this study was to indicate the spatial effect on wages by using coordinate variables called as longitude and latitude. The rest of the paper is organised as follows: Sections 2, 3 and 4 introduce Mincerian wage equation and literature, methodology and data used, respectively. Estimation results are presented and discussed in Section 5. The final section provides the conclusion.

2. MINCERIAN WAGE EQUATION AND LITERATURE

Most empirical studies of the relationship between education and earnings are based on version of Mincerian wage equation of schooling, experience and earnings. Mincer (1974) modelled the natural logarithm of earnings as the function of years of education and years of potential labour market experience. The standard mincerian wage equation can be formulated as follows:

$$W_i = \alpha_0 + \alpha_1 NS_i + \alpha_2 EXP_i + \alpha_3 EXP_i^2 + \varepsilon_i \qquad (1)$$

Where W is log of hourly wage, NS is the number of schooling years and EXP corresponds to Mincer experience was calculated as age minus schooling minus school starting age and ε is a random iid. disturbance term that reflects unobserved characteristics. In common usage, the coefficient on schooling in a regression of log earnings on years of schooling is often called a rate of return to education. This model also can include some other regressors that may affect earnings such as occupation, race, gender, maritus, region, race, number of children, firm size etc. Though lists of other regressors are typically added to the standard mincerian wage equation, the three key variables in equation still appear in most empirical estimates of earnings regressions (Guris & Caglayan, 2011). In the mostly used human capital model of Mincer (1974), logarithmic wages are modelled as a linear function of education of years and a quadratic function of years of experience. The studies based on standard Mincer wage equation added experience variable in quadratic form (Liu, 2004; Martins & Pereira, 2004; Nyhus & Pons, 2011; Tansel & Bodur, 2012; Mandel & Semyonov, 2014). However, Murphy and Welch (1990) examined several wage-experience profiles in their studies and observed that the estimation results are biased when the experience variable is added in a quadratic form into the model.

Murphy and Welch (1990) observed that while the wage model in quadratic form underestimated wages for individuals having 10 years experience, the initial wages are overestimated for individuals having any education level. As a result, Murphy and Welch (1990) stated that the wage model is the best model fitting the data set if the experience variable is specified as in a quartic form and supported this with empirical findings. Seltzer and Frank (2007) showed that the model in quartic form gives better estimation results than the model in quadratic form. Burnette and Stanfors (2012) analyzed that the data set fits best which model by adding the experience variable in quadratic, quartic and quadratic spline. According to this analysis, the model in quadratic form never fits the data set and causes to be biased estimation results. The model in quartic form is relatively better but the model in quadratic spline fits best the data set. The studies based on standard Mincer wage equation, age variable were added in a quadratic form such as experience variable (Aldashev et al., 2012; Azam, 2012; Furno, 2013; McNabb and Said, 2013). On the other hand, Hatton (1997) analyzed the model in which functional form fits better the data set by examining different wage-age profiles in his study. According to this study, while the model in quadratic form underestimates the wages for individuals in age between 20 and 30, it overestimates the wages for individulas in age between 30 and 50.

Hereupon, Hatton (1997) examined that the model fits best the data set by estimating wage models in different functional forms and concluded that the model in quadratic spline function form allows the parameters change before and after age 25 fits best the data set. Burnette and Stanfors (2012) made a comparison among different functional forms by creating wage and age profiles in order to determine the model fits best the data set. The model in quadratic spline form was determined as the best model fits the data set. For wage-age profiles estimated in quadratic spline form, age threshold values were determined between 17 and 28 and observed that wages how change before and after these threshold values. The age threshold value was determined as 20 age for male workers and 23 age for female workers. As a result, while in some studies wage models were estimated in different functional forms, in many studies the wage models were estimated in quadratic form as well as in standard Mincer wage equation. Accordingly, it should be investigated that whether Mincerian wage equation in quadratic form does or not fit good the data set before estimate the wage models.

3. METHODOLOGY

In a wage equation setting, parametric model can be written as follows:

$$E(LW \mid X) = \alpha + \beta' X \tag{2}$$

Where LW is the earnings or wages, X is the vector of the explanatory variables related to LW and β is a vector of coefficients. Such a model is an ad hoc parameterization of LW, which is quite restrictive. The present study attempts to extend previous analyses by enriching (2) with related variables which are non-parametrically related to LW. Denoting such a variable by Z, and incorporating it into (2) in a non-parametric manner yields the semi-parametric model in (3):

$$E(LW \mid X, Z) = \alpha + \beta' X + f(Z) \tag{3}$$

Where; $\alpha + \beta' X$ represents the parametric component, and f(Z) the non-parametric component. The non-parametric component f(Z) is estimated using splines with optimal basis functions, a method discussed analytically in Keele (2008). The logic behind a spline is to estimate separate regression lines that are joined at the corresponding knots. An important advantage of the splines methodology, in comparison to the commonly used piecewise regressions, is that it does not pre-specify ad hoc cut-off points. The employed methodology minimizes the following objective function:

$$\min\left\{\frac{1}{n}\sum_{i=1}^{n}(LW_i - f(Z_i) - \alpha - \beta' X)^2 + \lambda J\right\} \tag{4}$$

Where; J represents the roughness of function f and n denotes the number of observations. The previous expression describes the trade-off between fitting perfectly the data (i.e. minimizing the squared residuals) and having the smoothest possible approximating function f. This trade off is controlled by parameter λ. As $\lambda \to \infty$, the penalty assigned to the roughness of the function is so high that the optimal function, f, is of linear form, since, by definition, a linear function has zero roughness for the whole range of the dependent variable values. In this case, the minimization problem becomes identical to least squares. On the other extreme, if $\lambda \to 0$, then this methodology will provide a very rough approximating function f that essentially fits each individual observation. Instead of using smoothing splines as in Engle et al. (1986), this study employs penalized regression splines. Even though these two approaches yield similar results in practice, penalized regression splines use fewer parameters and are, therefore, computationally more efficient. Thus, the objective function becomes:

$$\min\left\{\frac{1}{n}\sum_{i=1}^{n}(LW_i - f(Z_i) - \alpha - \beta' X)^2 + \lambda\int f''(Z)d(Z)\right\} \tag{5}$$

Where f(Z) is a regression spline and f'' stands for the second derivative of f. The roughness of f(Z) is captured by

$$\int f''(Z_i)d(Z).$$

Model (3) can be easily generalized into the generalized additive model given in (6) which allows several explanatory variables, $Z_1, \ldots, Z_k$, to enter non-parametrically:

$$E(LW \mid X, Z) = \alpha + \beta' X + f_1(Z_1) + \ldots + f_k(Z_k) \tag{6}$$

This is essentially a penalized likelihood maximization problem solved by Penalized Iteratively Reweighted Least Squares (P-IRLS). The selection of the optimal smoothing parameter λ is integrated in this procedure using the Generalized Cross Validation (GCV) criterion. According to this criterion, the optimal λ minimizes the following expression:

$$GCV(\lambda) = \frac{RSS(\lambda)}{[1 - n - 1tr(S(\lambda))]^2} \tag{7}$$

Where; RSS(λ) = e'e is the sum of squared residuals of the estimated model for a given λ and tr(S(λ)) is the trace of the projection matrix S(λ) that satisfies $\hat{Q} = SQ$. For each of the models estimated, the corresponding minimized GCV scores are also reported (Eubank, 1999; Ruppert et al., 2003; Yatchew, 2003). Furthermore, this methodology enables us to test the statistical significance of each non-parametric variable in the specified semi-parametric model in (6). This is done via an F-test that compares the sum of squared residuals (RSS) of the semi-parametric model (unrestricted) with the RSS of the restricted model that excludes the non-parametric variable. The corresponding F statistic is given by:

$$F = \frac{(RSS_{restricted} - RSS_{unrestricted})/(tr(S) - 1)}{RSS_{unrestricted} / df_{res,unrestricted}} \tag{8}$$

Where; $df_{res} = n - tr(2S - SS')$. This statistic under the null hypothesis of equal RSS follows an F distribution with $df_{res,restricted}$-$df_{res,unrestricted}$ and $df_{res,unrestricted}$ degrees of freedom. Likelihood Ratio (LR) test can be used to test whether the semi-parametric model has explanatory power superior to that of the parametric model:

$$LR = -2 * (LogLikelihood_{restricted} - LogLikelihood_{unrestricted}) \tag{9}$$

This test compares the log-likelihood of the fully parametric model (restricted) with the loglikelihood of the semi-parametric model (unrestricted). The test statistic under the null hypothesis of equal likelihoods follows an approximate χ^2 distribution with degrees of freedom given by the difference in the number of parameters across the two models (Keele, 2008).

4. DATA AND VARIABLES

In this study, we examine that whether the standard mincerian wage equation is valid or not for private and public sector employees. Turkish Household Labour Force Survey data was used conducted by Turkish Statistical Institute in the year 2013. We divide the data two group and estimate the mincerian equation separately for both private and public employees for that year. The employees who did not work in the survey year were deleted. Also the employees between the ages 15 and 65 are considered. We modeled the natural logarithm of hourly wages (LW) as the function of education (EDC) defined as dummy variable, years of experience (EXP) was calculated as 2013 year minus year of starting to work, working hours (WH), age (AGE), coordinate variables named as longitude (LONG) and latitude (LAT), social insurance (SI), marital status (MS), permanency of job (PER), full time job (FULL), sector (SEC) and gender for full data set. In order to describe the variables in the models estimated for each group, we created Table1. In order to determine the distribution of wages, we report the descriptive statistics of the variables in sample in Table 2. The private and public average log wages were 6.810 and 7.510 in the year 2013, respectively. It can be seen that the private-public average wage differential was 0.7 for the year 2013. The experience difference between private and public sector employees was found 8 years in 2013. On the other words, employees working in the public sector have more experience than in private sector. Furthermore, while public sector employees' average age was 39, the average age was 34 in private sector. That is, younger employees were preferred in private sector.

5. EMPIRICAL FINDINGS

The relationships between wage and explanatory variables were investigated by semiparametric regression models without any restrictive assumption about functional form in order to avoid misspecification problem. In this section, semiparametric approach does not make any assumption about the functional form of the relationship between variables and determining the functional form by utilizing the distribution of data set will be introduced. As different from standard Mincerian wage equations, the wage model will not be estimated in quadratic form and the functional form of the model will be determined by utilizing distribution of data set. Therefore, while the quantative explanatory variables will be added nonparametrically into the model, the qualitative variables will be added parametrically.

Table 1. Variable Definition

Variable name	Definition
Dependent variable	
LW	natural logarithm of hourly wages
Non-parametric explanatory variables	
AGE	age
EXP	2013 year minus year of starting to work
WH	working hours
LONG	longitude
LAT	latitude
Parametric explanatory variables	
SI	social insurance ,dummy variable(1: if an employee has social insurance,or 0: if not)
MS	marital status, dummy variable (1: if married, or 0: if not)
GEN	gender, dummy variable (1: if male, or 0: if not)
PER	permanency of job, dummy variable (1: if job is permanent, or 0: if not)
FULL	full time job, dummy variable (1: if job is full time, or 0: if not)
PRIMARY	primary, dummy varibale (1: if education level is primary, or 0: if not)
SECONDARY	secondary, dummy varibale (1: if education level is secondary, or 0: if not)
HIGHSCHOOL	highschool, dummy varibale (1: if education level is highschool, or 0: if not)
UNIVERSITY	university, dummy varibale (1: if education level is university, or 0: if not)
SECTOR	sector, dummy variable (1: if private, or 0: if not)

Table 2 . Descriptive Statistics

	LW	EXP	WH	AGE
Private				
Mean	6.81	4.62	51.693	34.221
JB	16588.19	21010.97	961.569	330.416
Probability	0.000	0.000	0.000	0.000
Public				
Mean	7.51	13.435	40.842	39.457
JB	4579.416	171.449	9606.348	28.595
Probability	0.000	0.000	0.000	0.000

Coefficients of the dummy variables in parametric and semiparametric models, in semi-logarithmic form, were interpreted based on the approach exhibited by Halvorsen and Palmquist (1980). While the parametric model was estimated by Ordinary Least Squares (OLS), the semiparametric model was estimated by Penalized Iteratively Reweighted Least Squares (P-IRLS). The age, experience and working hours variables were added as quadratic form into parametric wage models. However, the effect of these variables on the wage can be different from by sector. For this reason, the relationships between wage and the explanatory variables were examined seperately based on private and public sector. Instead of apriori assumption about the functional form of the relationship, the functional form can be determined through the distribution of data set. For this aim, semiparametric model was estimated by adding subject variables in nonparametric form.

Moreover, the coordinates variables, including longitudes and latitudes of 26 subregions in Turkey, representing the spatial effect on the wages were also added in nonparametric form for semiparametric model. The subregions include provinces that working households dwell. In the presence of heteroscedastic errors in the linear model, the OLS estimator of the parameters is consistent but its variance is no longer so. Any tests based on these estimates will thus no longer will be valid. The same problem can arise in the parametric part of the semiparametric model. It is possible to calculate an estimate of the variance which is robust to heteroscedasticity, as proposed by White (1980) for both of models. For this reason, the standard errors of parametric and semiparametric model were obtained as robust to heteroscedaticity in all estimations. Moreover, Ramsey's RESET test designed to detect if there are any neglected nonlinearities in the parametric models was used. The empirical results of both the parametric and the semi-parametric models based on full data set, private/public data sets are reported respectively in Table 3 and Table 4. In the tables, t-values are given in parentheses. F-stat reports the F-test statistic value for the statistical significance of each non-parametric term in the semi-parametric model (the null hypothesis is that that each corresponding term is not statistically significant).

Table 3. Parametric and semi-parametric results, full data set

Variable	Parametric_full	Semiparametric_full
Intercept	3.326***	6.208**
	(26.84)	(149.815)
GENDER	0.179***	0.177***
	(20.22)	(20.275)
MS	0.069***	0.075***
	(6.57)	(7.101)
SI	0.304***	0.293***
	(26.53)	(25.657)
SECONDARY	0.087***	0.103***
	(6.80)	(8.065)
HIGHSCHOOL	0.207***	0.204***
	(18.36)	(18.305)
UNIVERSITY	0.666***	0.645***
	(54.52)	(52.963)
PERMANENCY	0.139***	0.127***
	(8.61)	(7.886)
FULL TIME	0.372***	0.213***
	(12.80)	(5.647)
SECTOR	-0.29***	-0.287***
	(-24.78)	(-23.646)
AGE	0.042***	See Fig.1
	(16.09)	F-stat: 42.51***
AGE2	-0.001***	-
	(-14.32)	
EXP	0.014***	See Fig.1
	(8.44)	F-stat: 73.11***
EXP2	-0.0002***	-
	(-2.75)	
WH	0.019***	See Fig.1
	(9.47)	F-stat: 22.67***
WH2	-0.0002***	-
	(-8.53)	
LONG	-0.004***	See Fig.1
	(-4.09)	F-stat: 37.91***
LAT	0.034***	See Fig.1
	(13.53)	F-stat: 25.59***
R^2 adjusted	59.3%	61.1%
LR test (statistic value)		590.068***
Ramsey RESET	131.340***	

Notes. Coefficient is statistically significant at the 1% level, 5% level and 10% level, respectively.

According to Table 3 and Table 4, it was observed that semiparametric form explains the wage equation better than parametric form. While the changes in explanatory variables explain 61.1 % (full data), 49.5 % (in private sector data) and 65.2 % (in public sector data) of changes in the wages for semiparametric model, the changes in explanatory variables explain 59.3 %, 46.9 % and 63.5 % of changes in the wages for parametric model, respectively. The results of LR test support also these findings and indicate that the explanatory power of the semiparametric model is better than the parametric model. Ramsey's RESET test shows that there are neglected nonlinearities in the parametric model. Table 3 includes the estimation results of the parametric and semiparametric model based on full data set.

In semiparametric model, the coefficients of the parametric variables can be interpreted by using Halvorsen and Palmquist approach. The effect of gender on wages is 19.36 %, that is male employees earn 19.36 % more than female employees. The effect of marital status on wages is 7.79 % and married employees earn 7.79 % more than unmarried employees. The employees having social insurance earn 34.04 % more than the others. The effect of education on wages increase gradually. Accordingly, the secondary education increases wages 10.85 % to others. The highschool education increases 22.63 %, while university education increases 90.59 %. If job is a full-time job, the wages increase 23.74 %. In private sector, wages are lower 24.95 % than public sector. For the semi-parametric models, the partial impact of each nonlinearly related variable on LW is depicted in Figure 1. The estimated coefficients of the variables added in nonparametric form into semiparametric or nonparametric models are demonstrated by figures. Since each observation in the data set is determined as target observation in selection of smoothing parameter for nonparametric estimations, the number of models are estimated as much as number of observations in the data set. Consequently, 12580 semiparametric wage models were estimated for all data set and 12580 estimated coefficients were obtained for each independent variable. Therefore, the estimated coefficients of the variables in nonparametric form are demonstrated by figures in Figure 1.

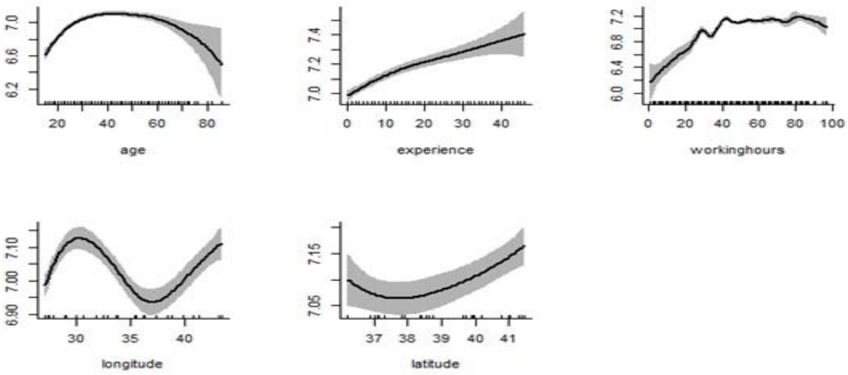

Figure 1.
Full Data Based Relations

The figure on age shows that the functional form of the relationship between wage and age is quadratic. At the beginning, the wages increase together with increment in age and achieve the maximum level at mid-forties. After this level, the wages decrease while the age increase. The experience on figure demonstrates that the wages increase slightly until two decades experience but after this level the wages increase acceleratingly. When the relationship between wage and working hours is investigated, the wages increase acceleratingly because of income effect while working hours increase at the begining. As the weekly working hours exceeded 40 hours, the wages do not increase by substitution effect. The figures on longitude and latitude variables were used to indicate the spatial effect on wages. According to these figures, Istanbul, Kocaeli-Gebze, Kars, Van-Ercis and Agri were determined as the regions which the wages are at the maximum level. In other words, wages are maximum in some provinces from the east of Turkey and Istanbul being a metropolitan. Because of development subsidy in the east of Turkey, wages can achieve the maximum level in these provinces. The regions that the wages are minimum determined as Malatya, Adiyaman and Gaziantep cities from the east and southern east of Turkey. The provinces having maximum wages are also shown on Turkey map in Figure 2. The vertical and horizontal red lines show respectively the longitudes and latitude.

Full data based relations map

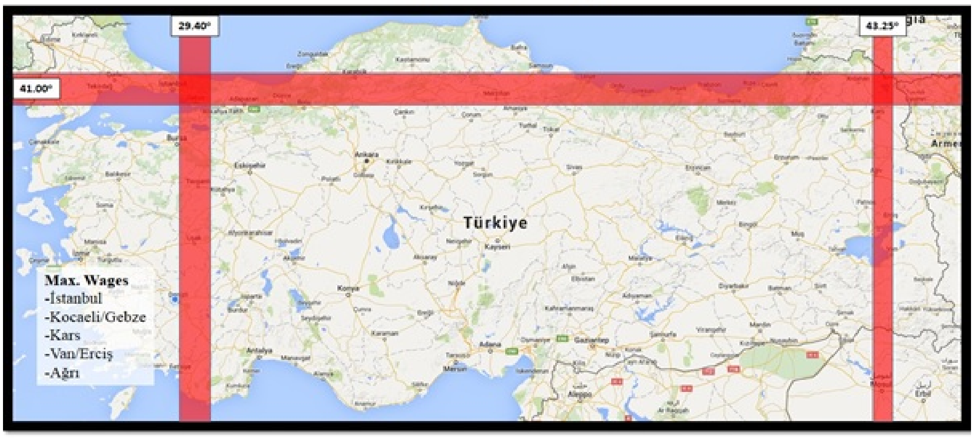

Figure 2.
Full Data Based Relations Map, Max Wages

The provinces having minimum wages are also shown on Turkey map in Figure 3. The vertical and horizontal black lines show respectively the longitude and latitude.

Full data based relations map

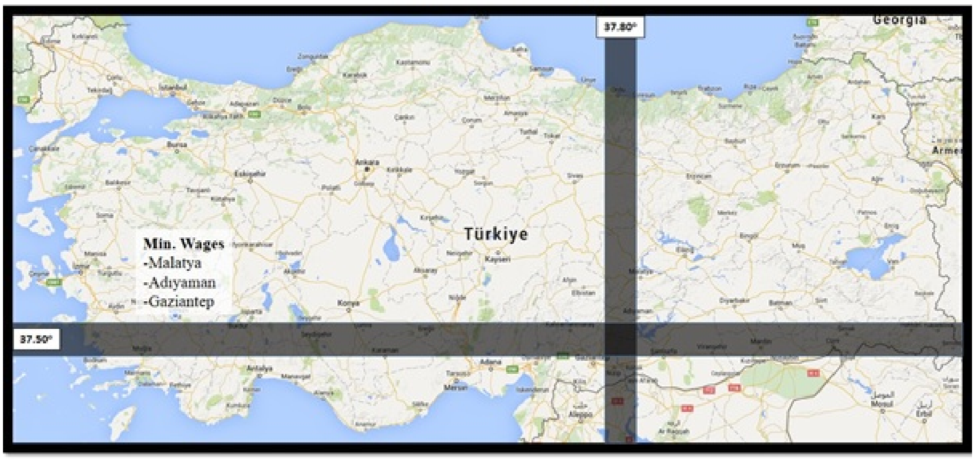

Figure 3.
Full Data Based Relations Map, Min Wages

The reason of minimum wages in these regions can be Syrian refugees immigrating to Turkey since the second half of the year 2011. The growing population of Syrian refugees in these counties and supplying their labour to the market in lower wages can be one of the reason of minimum wages. The estimation results of parametric and semiparametric wage models based on private and public employees are given in Table 4. The estimation results of semiparametric models in Table 4 can be interpreted comparatively for private and public employees. The effect of gender gap on wages is 21.16 % in private sector while the effect of gender gap on wages is 12.97 % in public sector. Married employees working in private sector earn 7.57 % more than unmarried employees while married employees working in public sector earn 5.02 % more than unmarried employees. While employees having social insurance in private sector earn 30.34 % more than the others, employees having social insurance in public sector earn 88.69 % more than the others. All education variables are statistically significant in semiparametric model based on private sector but the secondary variable is not statictically significant in semiparametric model based on public sector. The secondary educated employees earn 12.41 %, high school educated employees earn 21.16 % and university educated employees earn 89.26 % in private sector.

Table 4. Parametric and Semi-Parametric Results

Variable	Parametric_private	Semiparametric_private	Parametric_public	Semiparametric_public
Intercept	2.877***	6.121***	3.954***	5.836***
	(18.86)	(73.776)	(21.56)	(130.702)
GENDER	0.196***	0.192***	0.120***	0.122***
	(17.76)	(17.709)	(9.59)	(9.916)
MS	0.068***	0.073***	0.056***	0.049***
	(5.38)	(5.802)	(3.49)	(3.053)
SI	0.268***	0.265***	0.660***	0.635***
	(20.84)	(20.748)	(22.10)	(21.380)
SECONDARY	0.109***	0.117***	-	-
	(7.58)	(8.284)		
HIGHSCHOOL	0.191***	0.192***	0.275***	0.244***
	(14.93)	(15.146)	(13.72)	(12.301)
UNIVERSITY	0.667***	0.638***	0.587***	0.539***
	(44.44)	(42.530)	(32.57)	(29.571)
PERMANENCY	0.045***	0.046**	0.698***	0.585***
	(2.44)	(2.520)	(21.89)	(17.289)
FULL TIME	0.629***	0.073	-	-
	(14.97)	(0.863)		
AGE	0.043***		0.021***	See Fig.1
	(14.29)	F-stat: 37.03***	(4.32)	F-stat: 12.287***
AGE2	-0.001***	-	-0.0003***	-
	(-12.36)		(-4.27)	
EXP	0.014***		0.014***	See Fig.1
	(5.92)	F-stat:87.19***	(5.71)	F-stat: 53.151***
EXP2	-0.0001	-	-0.0002**	-
	(-1.11)		(-2.12)	
WH	0.013***		0.018***	See Fig.1
	(4.83)	F-stat: 13.79***	(7.36)	F-stat: 19.756***
WH2	-0.0001***	-	-0.00013***	-
	(-4.32)		(-5.43)	
LONG	-0.008***		0.004***	See Fig.1
	(-6.29)	F-stat: 31.70***	(3.60)	F-stat: 9.229***
LAT	0.041***		0.012**	See Fig.1
	(13.40)	F-stat: 23.36***	(3.32)	F-stat: 3.091**
R^2 adjusted	46.9%	49.5%	63.5%	65.2%
LR test (statistic value)		510.644***		164.235***
Ramsey RESET	4.734**		21.824***	

Notes. Coefficient is statistically significant at the 1% level, 5% level and 10% level, respectively.

In public sector, high school educated employees earn 27.63 % and university educated employees earn 71.42 % more than the others. Employees having a permanent job earn 4.71 % more in private sector while employees having a permanent job earn 79.49 % more in public sector. Finally, full time variable is not statistically significant in semiparametric models based on private and public sectors. The coefficient of squared experience is not statistically significant in semiparametric model estimated for private employees. However, the F test statistics show that all variables in nonparametric form are statistically significant for semiparametric model. If it is made a comparison between parametric and semiparametric models, it can be said that semiparametric form explains the wage equation better than the parametric form. The secondary and full time variables were not added into the parametric and semiparametric models because their coefficients are not statistically significant. For the semi-parametric models, the partial impact of each nonlinearly related variable on LW is depicted in Figure 4 and Figure 5. The estimated coefficients of nonparametric terms in semiparametric model for private employees are shown by figures in Figure 4.

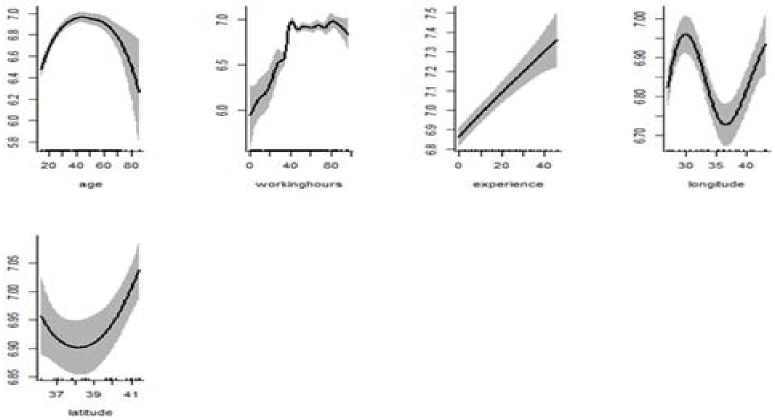

Figure 4.
Private Data Based Relations

The figure on experience shows that there is a linear relationship between wages and experience. This result shows that the wages increase while the experience increase. In this case if it is made an apriori assumption that the functional form of the relationship between wage and experience is quadratic, it would arise misspecification problem about the parametric wage model. The figure of age depicts a relationship in quadratic form between age and wages for private employees. The employees who are above 40 years of age are less prefered because of diminishing productivity and the wages can decrease. When the relationship between wage and working hours was investigated, it was observed that the wages increase acceleratingly because of income effect while working hours increase at the begining. As the weekly working hours exceeded 40 hours, the wages do not increase by substitution effect. On the other hand, the figures on longitude and latitude show that the regions which wages are minimum and maximum are same as full data based relations.

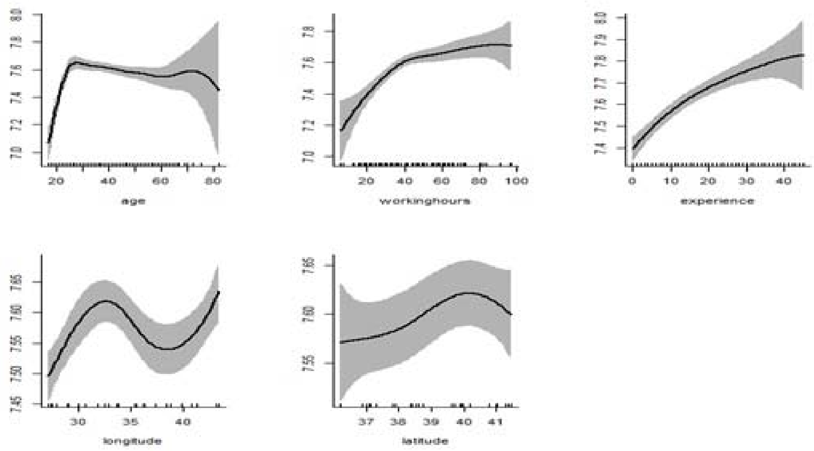

Figure 5.
Public Data Based Relations

The figures depicting the relationship between wages and nonparametric variables are given in Figure 5. Therefore, the figure on age shows that the wages increase rapidly till 30 age and then the wages decrease slowly. One of the reasons of this result can be that many employees working in the public sector getting promotion till in the range of 30-35 age and tendency of employees being at the same job position after 35 age. This result indicates in a sense that labor force in public sector is unproductive. Moreover, the age figure shows that there is not a relationship in quadratic form between wages and age for public employees. Thus, the functional form of the relationship between wages and age should not be determined as quadratic in parametric wage model for public employees. Otherwise, it would be a misspecification problem because of wrong functional form. The figure on working hours shows that the wages grow decreasingly after 40 weekly working hours.

When it is examined that the relationship between wages and experience, it is observed that the effect of experience on the wages in public sector increase slowly to private sector. This result can arise from lower wage dividend by experience in public sector compared to private sector. The figures on longitude and latitude demonstrate the spatial effect on wages in public sector. According to these figures, the wages are maximum in Ankara and nearby cities, Kars, Van-Ercis and Agri for public sector. This can result from development subsidy taken by public employees working in the east of Turkey. On the other hand, the provinces having minimum wages can be implied as Izmir, Canakkale, Diyarbakir and Sanliurfa. The provinces having maximum wages can be also shown on Turkey map in Figure 6. The vertical and horizontal red lines show respectively the longitudes and latitude.

Figure 6.
Public Data Based Relationships Map, Max Wages

The provinces having minimum wages can be also shown on Turkey map in Figure 7. The vertical and horizontal black lines show respectively the longitudes and latitude.

Figure 7.
Public Data Based Relationships Map, Min Wages

6. CONCLUSION

The Mincerian wage equation is one of the most widely used among the human capital models. In the mostly used human capital model of Mincer (1974), logarithmic wages are modelled as a linear function of education of years and a quadratic function of years of experience. This model also can include some other regressors that may affect earnings such as occupation, race, gender, maritus, region, race, number of children, firm size etc. Though a list of other regressor is typically added to the standard mincerian wage equation, age, education and experience variables in equation appear in most empirical estimates of earnings regressions. In many studies, these variables are also added in quadratic or quartic form into the Mincerian wage equation. Therefore, it should be investigated that whether Mincerian wage equation in quadratic form does or not fit good the data set before estimate the wage models. The aim of this study was to find out the best functional form of mincerian wage equation in Turkey. For this aim, semiparametric regression model determining the functional form through distribution of data set without any restrictive assumptions about functional forms between variables was used. Moreover, in order to make comparison between parametric and semiparametric models, parametric wage models were also estimated. Besides the key variables in mincerian wage equation, social insurance, marital status, sector, full time job, permanency of job and the coordinate variables measuring the spatial effect on wage were added into the model. After estimation of semiparametric regression models, different empirical profiles were obtained based on male, female, private, public and full data set. The data used was obtained from Household Labour Survey in 2013.

The findings of the study show that there is no quadratic relationship between wage and experience for private employees by contrast with standard mincerian wage equation. However, there is a linear relationship for private employees. The linear relationships between wage and experience show that the wage dividend is higher by experience in private sector. The age variable is added in quadratic form into wage models in many empirical studies. However, it was found that there is not a quadratic relationship between wage and age for public employees in this study. Therefore, the figure on age shows that the wages increase rapidly till 30 age and keep up with for a while then wages decrease slowly. One of the reasons of this result can be that many employees working in the public sector getting promotion till in the range of 30-35 age and tendency of employees being at the same job position after 35 age. This result indicates in a sense that labor force in public sector is unproductive.

The study provides some evidence that wages are maximum in Istanbul and nearby cities with some provinces from east of Turkey. The development subsidy can be one of the reason of maximum wages in some provinces from the eastern Anatolia region. The provinces that the wages are minimum determined as Malatya, Adiyaman and Gaziantep. The reason of minimum wages can be the growing population of Syrian refugees since second half of the year 2011 and supplying their labour to the market in lower than Turkish employees. These results are valid for full and private data apart from public data. The wages are maximum in Ankara and nearby cities, Kars, Van-Ercis and Agri for public sector. As mentioned before, this can result from development subsidy taken by public employees working in the east of Turkey. On the other hand, the provinces having minimum wages can be implied as Izmir, Canakkale, Diyarbakir and Sanliurfa. Moreover, some remarkable estimation results were obtained from semiparametric models. The effect of education on wages in private sector is more than public sector and male employees earn more than female employees in private sector.

Consequently, the relationship between wages and the key explanatory variables in the standard mincerian wage equation are not always in quadratic form. The functional forms between wage and subject variables can be different from quadratic form and these differences can arise from gender, sector and region. For this reason, the functional form of the relationship between wages and explanatory variable should not be assumed apriori in quadratic form and investigated firstly their empirical profiles. In this case, the semiparametric regression or nonparametric model do not make apriori assumptions related to the functional forms can be used. Otherwise, the regression models will be misspecified.

REFERENCES

Aldashev, A., Gernandt, J. and Thomsen, S. (2009), The Immigrant Wage Gap in Germany. Centre for European Economic Research (ZEW).

Azam, M. (2012), Changes in Wage Structure in Urban India, 1983–2004: A Quantile Regression Decomposition. World Development, 40(6), 1135–1150.

Burnette, J. and Stanfors M. (2012), Before Careers: Experiences of Wage Growth among Late Nineteenth-Century Swedish Cigar Workers. Lund Papers in Economic History, 121.

Engle, R. F., Granger, C.W.J, Rice, J. and Weiss, A. (1986), Semiparametric Estimates of The Relation between Weather and Electricity Sales. Journal of the American Statistical Association,81, 310–320.

Eubank, R. L. (1999), Nonparametric Regression and Spline Smoothing, Statistics: Textbooks And Monographs. Second Edition, Marcel Dekker, Inc.

Furno, M. (2013), Quantile Regression and Structural Change In The Italian Wage Equation. Economic Modelling, 30, 420–434.

Güris, S. and Caglayan, E. (2012), Returns to Education and Wages in Turkey: Robust and Resistant Regression. QualityandQuantity, 46:1407-1416.

Hatton, T. J. (1997), The Immigrant Assimilation Puzzle in Late Nineteenth Century America. Journal of Economic History, 57, 34-62.

Keele, L. (2008), Semiparametric Regression for the Social Sciences. John Wiley and Sons, Ltd, England.

Liu, A. Y. C. (2004), Sectoral Gender Wage Gap in Vietnam. Oxford Development Studies, Vol. 32, No. 2.

Mandel, H. and Semyonov, M. (2014), Gender Pay Gap and Employment Sector: Sources of Earnings Disparities in the United States, 1970–2010. Demography, 51, 1597–1618.

Martins, P. S. and Pereira, P. T. (2004), Does Education Reduce Wage Inequality? Quantile Regression Evidence From 16 Countries. Labour Economics, 11, 355– 371.

McNabb, R. and Said, R. (2013), Trade Openness and Wage Inequality: Evidence for Malaysia. The Journal of Development Studies, 49(8), 1118-1132.

Mincer, J. (1974), Schooling, Experience and Earnings. Columbia University Press, New York.

Murphy, K. M. and Welch, F. (1990), Empirical Age-Earnings Profiles. Journal of Labor Economics, 8 (2), 202-209.

Nyhus, E. K. and Pons, E. (2011), Personality and The Gender Wage Gap. Applied Economics, 44, 105–118.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003), Semiparametric Regression, Cambridge University Press.

Seltzer, A. and Frank, J. (2007), Promotion Tournaments and White Collar Careers: Evidence from Williams Deacon's Bank, 1890-1941. Oxford Economic Papers, 59, 49-72.

Tansel A. and Bodur, B. F. (2012), Wage Inequality and Returns to Education in Turkey: A Quantile Regression Analysis. Review of Development Economics, 16(1), 107–121.

Yatchew, A. (2003), Semiparametric Regression for the Applied Econometrician, Cambridge University Press.