RELATION BETWEEN HEALTH AND WAGES IN TURKEY

Sezer Alcan, Ph.D.
Ministry of Treasury and Finance
Inonu Bulvari No 36
06510 Emek Ankara Turkey
sezer.alcan@hazine.gov.tr

Onur Özsoy, Ph.D.[1]
SoLBridge International School of Business
128, Uam-ro (151-13, Samsung-dong)
Dong-gu, Daejeon 34613, Republic of Korea
ozsoyonur@gmail.com

Received: 20 January 2017; Accepted: 06 February 2019.

Abstract

The purpose of this study is to estimate the effects of health on the hourly wages of women and men in Turkey by using panel data. The data are used to estimate the earning function, where the natural logarithm of hourly wage is the function of individual characteristics, including health. This work complements previous studies by using a panel in which the education variable, measured by the degree obtained, varies over time and therefore it can be estimated through the within estimator. One of the most important observations of this study is that very good and/or good self-assessed health status has a positive effect on wages more for women than for men. Another important finding is that of significant difference in the rate of return to education, which is higher for women than for men.

Keywords: Self-assessed health, health limitation, wages, panel data, fixed effects, Turkey.
JEL: J01.

1. Introduction

In the model of Michael Grossman (1972), health is an important element of human capital, and investment in health increases productivity and the number of working hours. Specifically, the relationship between health and work is much more noticeable for developing countries than for developed countries because the workable population in developing countries does not get adequate nutrition and has poor health. In addition, the efficiency wage theory provides a starting point for studies focusing on the outcome of health and labor markets in developing countries (Janet Currie and Brigitte C. Madrian; 1999, Harvey Leibenstein, 1957; Partha Dasgupta, 1997). For developed countries, an increase in health conditions improves productivity and hence the wage rate (Selma Mushkin, 1962, Harold S. Luft, 1975).

The purpose of this study is to estimate the effect of health on the hourly wages of women and men in Turkey by using panel data. In most previous studies, the education variable is time-invariant, and therefore its coefficient cannot be estimated through the within estimator. This work complements earlier studies by using a panel in which the education variable, measured by the degree obtained, such as a diploma, varies over time. The use such data eliminates the need for instrumental variables techniques, which have been the subject of criticism.

The report is organized as follows. Section 2 reviews relevant earlier studies on health and wages. Section 3 explains the estimation methods. In Section 4, the data set is presented, and the variables used in the model are described. The estimation results are presented and evaluated in Section 5. Section 6 concludes the analysis.

2. Literature Review

In his seminal work on human capital, Gary S. Becker (1962) stated that one way to invest in human capital was to improve health, in addition to other factors, such as schooling. Becker's model was further developed by Grossman (1972). The Grossman (1972) model regards health as a part of human capital that produces healthy time and therefore increases the number of working hours and productivity.

The existing literature on health and labor productivity is mainly on developing countries. Because the labor force in developing countries is observed to be undernourished and in bad health, the efficiency wage theory applies (Currie and Madrian; 1999, Leibenstein, 1957; Dasgupta, 1997). Studies that use data from developing countries mainly apply nutritional status and anthropometric measurements, such as weight, height, and body mass index (BMI), as health variables. On the other hand, studies on developed countries mostly use self-reported health status and presence of chronic conditions (Currie and Madrian, 1999; John Strauss and Duncan Thomas, 1998; Duncan Thomas and Elizabeth

[1]Corresponding author.

Frankenberg, 2002).

For example, the study by Thomas and Strauss (1997) used cross-sectional data to analyze the impact of health indicators on the wages of men and women in urban Brazil. The health variables used in the analysis were body mass index (BMI), height, and protein intake. The study found that height had a substantial effect on wages and that BMI had a positive impact on the wages of men, specifically among those less educated.

One of the first studies on this topic was carried out by Lung-Fei Lee (1982). Based on the generalized version of the model of James J. Heckman (1978), Lee used a cross-sectional sample of male US citizens to assess the impact of health on wages. Health was measured by self-assessed health and functional limitation variables. According to the analysis, health had a positive impact on wages, and vice versa. In another study, Lixin Cai (2007a) confirmed that health positively affects labor force participation. Cai (2007b) estimated a multi-equation system using cross-sectional Australian data and found that health, measured by self-reported health status, had a positive impact on wages when endogeneity was considered.

In another study, Paul Contoyannis and Nigel Rice (2001) used the British household panel survey and estimated the earnings function for males and females. However, in their data set, the education variable was time-invariant and could not be estimated by the fixed effects method; therefore, they applied the instrumental variable method proposed by Jerry A. Hausman and William E. Taylor (1981), Takeshi Amemiya and Thomas E. MaCurdy (1986), and Trevor Breusch, Grayham Mizon, and Peter Schmidt (1989). The results indicated that poor psychologic health status decreased the hourly wages for males, whereas excellent self-assessed health status improved hourly wages for females.

In a more recent study, Robert Jäckle and Oliver Himmler (2010) applied the method proposed by Semykina and Wooldridge (2010). They used German panel data and found evidence that selection correction was necessary. According to their analysis, good health status induced higher wages for men; however, no such effect was observed for women. The method suggested by Anastasia Semykina and Jeffrey M. Wooldridge (2010) requires a balanced panel. However, the way Jäckle and Himmler (2010) coded missing values for the variables of interest raised a question as to whether a balanced panel approach and selection correction were actually applied in their study. For example, for occupational classes, they created a new variable labeled as “missing occupation” for cases in which the occupational class was missing. Therefore, they coded missing values as if they were different variables.

3. Estimation methods

In the labor economics literature, the model of Jacob Mincer (1958, 1974) has been extensively used in the empirical estimation of the earnings function. The model shows how labor market rewards qualifications, such as experience and education, which has apparently direct impact on productivity (James Heckman, Lance J. Lochner and Petra E. Todd, 2003). The standard model has been amended considering the panel characteristic of data and to determine the impact of health and other variables on wages. The resulting model can be expressed as follows:

$$w_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \qquad t = t_i, \dots . T_i \qquad (1)$$

where i represents the individual and t denotes the time period. The data set is an unbalanced panel; therefore, some individuals do not appear in all time periods. In the first expression, w_{it} is the logarithm of hourly wages; x_{it} is a 1xK vector of explanatory variables such as experience, health, and marital status; α_i is the unobserved heterogeneity; and ε_{it} is the error term. If the unobserved heterogeneity and the error term are collected and expressed as the composite error term $v_{it} \equiv \alpha_i + \varepsilon_{it}$, Eq. (1) can be written as:

$$w_{it} = x_{it}\beta + v_{it} \qquad (2)$$

If there is no correlation between the composite error term and the explanatory variables, that is $(x'_{it}v_{it}) = 0$, $t = t_i, \dots . T_i$, the *pooled ordinary least square* estimation of Eq. (2) is consistent (Wooldridge, 2002). Practically, the absence of correlation between, x_{it} and v_{it} requires two important assumptions; (i) $E(x'_{it}\varepsilon_{it}) = 0$ and (ii) $E(x'_{it}\alpha_i) = 0$. If $E(w_{it}|x_{it}, \alpha_i)$ is successfully modeled, $E(x'_{it}\varepsilon_{it}) = 0$ holds. However, even if the assumption of $E(x'_{it}\alpha_i) = 0$, holds, the composite error term will be serially correlated, because there will be an unobserved effect, α_i, in each observed time period. Therefore, a robust variance matrix estimator and robust test statistics are necessary for pooled OLS (Wooldridge, 2002). Estimation by pooled OLS assumes α to be constant, similarly to a cross-section analysis. If this assumption is correct and the error term is not correlated with the explanatory variables, OLS will achieve consistent and unbiased estimates. However, OLS does not use the panel structure of the data (Wooldridge, 2002).

Therefore, when the panel characteristics of the data are applied, the random effects (RE) model is used. This model requires strict exogeneity, along with orthogonality, between α_i and x_{it}. The strict exogeneity assumption is expressed as $E(v_{it}|x_{it}, \alpha_i) = 0$, and the orthogonality assumption as $E(\alpha_i|x_{it}) = E(\alpha_i)$. Another assumption that is required for the RE model is the rank condition: $E(X'_t\Omega^{-1}X_i) = K$, where Ω is unconditional variance matrix of v_i, $\Omega \equiv E(\varepsilon_i\varepsilon'_i)$. The efficiency of the RE necessitates $E(\varepsilon_i\varepsilon'_i|x_i, \alpha_i) = \sigma_u^2 I_T$, and $E\left(\alpha_i^2\middle|x_i\right) = \sigma_\alpha^2$. When these conditions are met the RE method is asymptotically equal to the generalized least squares (GLS) technique. Because the data are drawn from a large population, the use of the RE model seems reasonable. However, if the random error for each cross-sectional unit is correlated with any of the other explanatory variables, the RE model will provide biased estimates (Wooldridge, 2002).

Assuming that the unobserved heterogeneity α_i, and the explanatory variables x_{it}, are correlated, the fixed effects (FE) method is used. The first assumption for the FE model is strict exogeneity: $E(\varepsilon_{it}|x_i, \alpha_i) = 0$, $t = t_i, \dots, T_i$. Under this assumption, the estimates obtained from the FE method are unbiased. The orthogonality assumption is not needed in the FE model, in contrast to the RE model. The FE estimator is more robust than the RE estimator. However, this requires that the explanatory variables do not include any time constant components. The second assumption for the FE model is the rank condition; $\sum_{t=1}^{T}(\ddot{x}'_{it}\, x_{it}) = K$ where $\ddot{x}'_{it} = x_{it} - \bar{x}$. This assumption ensures that there is no multicollinearity. The third assumption, $E(\varepsilon_i \varepsilon'_i | x_i, \alpha_i) = \sigma_\varepsilon^2 I_T$, implies that the variance of ε_i is constant across all periods and are serially uncorrelated. This assumption guarantees the efficiency of the FE estimates (Wooldridge, 2002). The major drawback of the FE model is that time constant components cannot be included in the explanatory variables. This has been a problem for most of the earlier studies in the literature because education could not be included as a regressor. However, this is not a concern in the present work, in which the education variable varies for some individuals.

4. Data set and definition of variables

4.1 Sample Construction

The Turkish Statistical Institute (TurkStat) has been carrying out the Income and Living Conditions Survey (TILCS) since 2006. This survey aims to provide data comparable with those in EU countries. Therefore, besides the national conditions, the standards of the European Statistical Office (EUROSTAT) were considered in designing the survey.

A rotational design is used in the panel survey. It is anticipated that 25% of the sampling size has been foreseen to get out of the frame of the panel from one year to another. Individuals 13 years and older from the selected basic sample of households are included in the sample and monitored over a period of 4 years.

The present analysis uses individual-level data from the TILCS for the years 2007 to 2011. The sample constructed from the survey consists of employed adults ages 18 to 66 years. Individuals without any formal education were dropped from the sample. Because this study aims to investigate productivity in terms of hourly wages, public employees, including military personnel, were dropped from the data set because their salaries are determined on a yearly basis by the government. Similarly, entrepreneurs and self-employed individuals were dropped from the sample. The sample includes only those individuals who responded to the question on earnings, thus allowing the calculation of hourly wages.

The study uses an unbalanced panel. Balanced panels include observations for all time periods for the same individuals, which helps control individual heterogeneity. However, when the data obtained from the TILCS were balanced for the variables of interest, such as wages, health, education, and occupational classes, the sample size was considerably diminished. In addition, the balanced panel constructed for the variables of interest may not represent the whole population, because it shows that individuals, especially women, have a very high educational attainment and are almost all employed in professional occupations with social security coverage. However, it is well observed that this was not the case in Turkey during the study period. Besides, the original, unbalanced panel shows quite the opposite. This fact supports the use of unbalanced data.

Labor force participation behavior has been observed to differ by gender. Therefore, the analysis is carried out separately for men and women. In particular, the sample includes those individuals who gave responses from which hourly wages could be calculated. This sample includes 5176 men and 3365 women on average.

4.2 Definition of Variables

4.2.1 Dependent Variable

To examine the impact of health on labor productivity, the logarithmic hourly wage of the individual is used as a dependent variable. The TILCS includes the weekly hours worked and net annual earnings of the respondents. The reference period for the income variable is "the previous calendar year.” Thus, the income declared in 2011 refers to the total income earned in 2010. The reference period for labor information is the previous week from the survey and the current date. The hourly wage is calculated by dividing the annual net payment by 52, the number of weeks in a year, and then by the weekly hours worked. The average hourly wage is 4.29 TL for males and 4.51 TL for females. Interestingly, women have approximately 5% higher hourly wages than men; previous studies on developed countries have reported the opposite.

Table 1 provides the definitions of the variables. The analysis of the mean of the occupational classes indicates that 22% of women work in professional occupations, compared with 9% of men. Whereas 44% of women work in skilled nonmanual occupations (such as associate professionals, clerks, service workers, and shop and market sales workers), 40% of men work in skilled manual occupations (such as skilled agricultural and fishery workers, craft and related trade workers, and plant and machine operators). This is better understood when labor force participation and the number of individuals declaring earnings are examined. Based on the sample, 19% of women have worked for pay in the past week, but 71% have declared wage incomes from last year. In contrast, 62% of men have worked for pay in the past week, and all of them declared wage incomes from the last year. These ratios clearly show that the labor force participation of women in Turkey is much lower than that of men; however, the women participating in the labor force have higher-paying occupations.

4.2.2 Explanatory Variables

Most of the previous studies on the impact of health on wages in developed countries used self-reported general health status, functional limitations or chronic conditions, and rarely, clinical assessments as health measures. Studies on developing countries, on the other hand, used nutrition status and/or weight, height, and body mass index, among others, as health variables. This is due to the fact that the association between health and labor productivity is quite noticeable in developing countries, in which the labor force is observed to be undernourished and in poor health. The theory of efficiency wages provides a good framework for the examination of this topic. In the present study, three measures of health are used. The first is *the self-assessed health* variable. In the TILCS, individuals are asked to rate their general health. The possible answers to this question are excellent, good, fair, poor, and very poor. Dummy variables were designated as equal to 1 if an individual has excellent health, good or fair health, or poor health, and 0 otherwise. Because the proportion of individuals who report having excellent health is quite low, excellent and good health were combined to represent one dummy.

The second health variable used in this study is *functional, physical, and psychological limitation*. In the TILCS, people are asked if their daily activities are limited due to a *physical and/or psychological problem* they had in the past six months. There are three possible answers to this question: "*yes, it was limited very much*," "*it was limited*," and "*no, it was not limited.*" A dummy variable equal to 1 was generated for each answer. Because the ratio of individuals giving the first response is low, the dummy variables for the first and second questions were combined to represent *physical and/or psychological problem limiting daily life.*

The third health variable used is *nutrition.* This variable comes from the household part of the survey. In the TILCS, households are asked whether they could eat meat, poultry or fish every two days (equivalent food for vegetarians). A dummy variable equal to one was created for "yes" responses. In addition, another variable was generated by multiplying the nutrition dummy with individuals working in unskilled jobs to measure the impact of nutrition on unskilled workers.

Education is another explanatory variable included in the model for analysis. Highest academic qualification obtained is used as a measure of educational attainment. In contrast to most of the earlier studies, the education variable in the data set used in this study is time variant; for some workers, there is a transition in educational attainment level from secondary school to high school or from high school to university.

Experience, which is the number of years the individual has been working, is included in the analysis in two ways: the level and its square. Due to the possibility of multicollinearity, age is excluded from the analysis. Dummy variables representing the occupational classes are also included in the estimation. The sample used in the analysis proportionally includes more women than men who are employed in professional occupations and skilled nonmanual occupations, such as associate professionals, clerks, service workers, and shop and market sales workers. Three dummy variables representing the sector in which a person works are also included in the model. These are manufacturing, construction, and wholesale/retail sectors. A variable representing the firm size, measured by the number of employees, is also included in the model. Another dummy variable, which indicates if a person is married, is used as well.

In the TILCS, individuals are asked whether they have social security coverage from their employers. In the sample, 44% of men work under social security coverage from their employers, compared with 14% of women. A dummy variable indicating social security coverage is included in the model. The TILCS also includes a question on the contract type of the worker, with three possible responses: a permanent employment contract, a fixed-term employment contract, and a temporary employment contract. A variable indicating a permanent employment contract is included in the model.

<<<<<Table 1 insert here>>>>>>>>>>

<<<<<<<Table 2 insert here>>>>>>>>>>>>>>>

5. Estimation results

5.1 Males

Table 3 presents the results for males. The sample includes the hourly wages of the men. To ensure the absence of heteroscedasticity and serial correlation, cluster-robust estimation technique is implemented, as suggested by Wooldridge (2002).

The estimation technique is described in the following. As a first step, the error term and explanatory variables are assumed to be uncorrelated, and thus the model is estimated by using OLS. This assumption guarantees unbiased and consistent estimates by OLS. On the other hand, OLS gives inefficient parameter estimates because it does not consider the use of panel data.

From the OLS results, the coefficient for self-assessed health is found to show the anticipated positive sign. The coefficient for very good and/or good self-assessed health is significant at the 5% level. The estimated coefficient for functional limitation, that is, psychologic or physical problem affecting daily life, is negative as expected and significant at the 1% level. The coefficient for nutrition, measured by protein intake, shows the expected positive sign and is significant at the 1% significance level. The coefficient for nutriunskll, the variable indicating the effect of nutrition on an individual with an unskilled job, has the expected negative sign; the estimated coefficient for that variable is not statistically significant.

OLS does not consider the panel nature of the data; thus, the model is estimated by using the random effects (RE). The estimated coefficient for the self-assessed health variable obtained by the RE is almost the same as the OLS estimate and remains significant at the 5% level. The estimated coefficient for functional limitation, which has a negative coefficient, is nonsignificant under the RE estimation. The coefficient for nutrition remains significant at the 1% level. The coefficient for nutriunskll is not significant and retains the expected negative sign. The coefficients for education variables show that, compared with technical school, university degree has a higher association with higher wages. The coefficients for both education variables have the expected positive signs and are significant at the 1% level. The coefficients for occupational status clearly show a pattern with increased wages; as an individual moves from a skilled manual work type (agrfish, crafttrade, machop) to a skilled nonmanual work type (ascprof, clerk, servshop) and from managerial to professional occupations, the hourly wages increase. The estimated coefficients for experience and its squared form have the anticipated concave property with the logarithm of hourly wages: as experience increases, so do wages; however, after some point the number of years the individual has been working negatively affects hourly wages. The estimated coefficient for the variable indicating marital status as married suggests that married individuals tend to have higher wages than unmarried ones. Employees who work in large organizations seem to have higher wages; however, the coefficients for firm size are insignificant under the RE estimation. Additionally, it is important to mention that the variable indicating individuals who have a permanent employment contract have higher wages.

By relaxing the assumption that the regressors are not correlated with the unobserved heterogeneity, α_i, the model can be estimated by using the fixed effects. The key advantage of using the fixed effects model is that the estimator remains consistent even when there is a correlation. In addition, because the education variables in the data set are time-variant, the effect of the extra education on wages can be measured by using the fixed effects.

The FE estimates of the health variables are lower compared with the RE estimates and retain their expected signs. However, the coefficients for both good self-assessed health and functional limitation variables are insignificant under the FE estimation. Nevertheless, the coefficient for nutrition remains significant at the 1% level. The coefficient for nutriunskll remains insignificant and retains the expected negative sign. Additionally, the estimated coefficients for the education variables are crucial. The results show that university degrees, compared with technical school degrees, are associated with higher wages. Nonetheless, the coefficients obtained from the FE estimation are lower than those from the RE estimation.

The coefficients for occupational status are lower than those obtained from the RE but can be interpreted similarly. The estimated coefficients for experience and their squared form have the anticipated concave function with the logarithm of hourly wages, and these estimates are almost the same as those obtained from RE. The estimated coefficient for the variable indicating marital status as married suggests that married individuals tend to have higher wages. The estimated coefficient for the variable indicating the social security coverage of individuals is significant at the 5% level but lower than the RE estimate.

There are differences between the results obtained from the FE and RE procedures. Assuming that the correct model specification was used and there was no correlation between the individual effects and regressors, the results from the FE estimation should be similar to those obtained by the RE procedure. Hence, the difference between the parameters obtained from the FE and RE methods can be formally tested. According to the standard Hausman test (Hausman, 1978) ($\chi^2_{25} = 480.78,\ Prob > \chi^2_{25} = 0.000$) FE is the true model.

A limitation of of the standard Hausman test is that it requires the RE estimator to be efficient. Therefore α_i and ε_{it} should be independently and identically distributed (iid). If α_i and ε_{it} are not iid, then the random effects estimator is not fully efficient under the null hypothesis of $E(\alpha_i + \varepsilon_{it}|x_{it}) = 0$. Wooldridge (2002) proposed a Wald test based on cluster-robust standard errors. The robust Hausman test gives similar test statistics ($\chi^2_{25} = 314.464, \ \ Prob > \chi^2_{25} = 0.000$) and confirms FE as the true model.

<<<<<<<<<<**Table 3 insert here**>>>>>>>>>>>>>>>>

5.2 Females

Table 4 shows the results for females. The FE estimates of the self-assessed health variable are higher than the RE estimates and are significant at the 1% level. Regardless of the weekly hours worked, the coefficient for the self-assessed health variable obtained from the FE is almost 2 or 2.5 times higher for females than for males. The estimated coefficient for functional limitation is positive but insignificant under both the FE and the RE estimation. The coefficient for nutrition shows the expected positive sign but is insignificant under the FE estimation.

In addition, regardless of the estimation method applied, the estimated coefficients for both education variables, technical high school degree and university degree, are significant at the 5% level. A significant feature of the coefficients for education variables needs to be addressed. The coefficients for education variables obtained from the FE are considerably higher than those obtained from the RE. The FE estimate for technical high school degree is almost thrice higher than the RE estimate, whereas the FE estimate for university degree is 28% higher than the RE estimate. This result is notable when compared with the findings in the male sample. Under the FE estimation, the rate of return to technical high school degree is 0.149 for males and 0.514 for females. The estimated coefficients for university degree under the FE estimation are much more striking: the rate of return to a university degree is 0.606 for females and 0.207 for males. However, the estimated coefficients for the education variables do not differ between males and females under the RE estimation. This indicates evidence of selection to market employment by educated women. However, the present FE results cannot be compared with previous findings because education is time-invariant in most studies and thus cannot be estimated using the FE method. This is the reason why most studies that use panel data receive criticism regarding their use of instrumental variable methods. For example, Contoyannis and Rice (2001) obtained a rate of return to a degree of between 0.7 and 1.2 by using instrumental variable methods, compared with around 0.4 under the RE estimation.

There are interesting differences in the results obtained from the male and the female data sets for the other variables. For example, work experience is significant at the 5% level for both males and females; however, each year of experience is rewarded more for women than for men. Also, the estimated coefficient for permanent employment contract is negative for females under the FE estimation, which is contrary to the expectation.

In light of the differences between the results obtained from the FE and RE estimations, the standard Hausman test is carried out to test the variances between the parameters. The standard Hausman test ($\chi^2_{25} = 147.67$, $Prob > \chi^2_{25} = 0.000$) shows FE as the true model. A robust Hausman test proposed by Wooldridge (2002) is applied and the test statistic confirmed the standard Hausman test result ($\chi^2_{25} = 108.874$, $Prob > \chi^2_{25} = 0.000$).

<<<<<<<<Table 4 insert here>>>>>>>>>>>>>>>

6. Conclusions

The aim of this study was to estimate the impact of health on hourly wages by using a Turkish panel data. The analysis is based on individual-level data obtained from 4 years of the TILCS. The sample consisted of employed adults ages 18 to 66 years, excluding public employees and self-employed labor. The data were used in the estimation of the Mincer-type earnings function; the natural logarithm of hourly wages was defined as the functions of health, work experience, academic attainment (degree), occupation, social security, firm size, and job permanency. The analysis was carried out separately for men and women. Self-assessed health status, psychologic or physical health limitation, and adequate nutrition measured by protein intake were used as health indicators. In earlier studies, education was a time-invariant variable, and its coefficient could not be obtained by the FE estimator. This study complements previous works by using a panel in which the education variable, measured by degree obtained, varies over time. The use of such data eliminates the need for the often criticized instrumental variables techniques.

The estimates obtained prove that health has an impact on the hourly wages of the individuals in the sample. Very good and/or good self-assessed health is observed to positively affect wages more for women than for men. Another interesting finding of this study is the difference between women and men regarding return to education. According to the FE estimation, the rate of return to education is higher for women than for men. This result indicates evidence of selection to market employment by educated women. The analysis focused on the impact of health variation on employed labor. This study could have been improved by considering selective participation in the labor force. However, due to the unbalanced nature of the panel used in this analysis, the sample selection problem could not be addressed.

.

Tables and Figures

Table 1: *Definitionsof Variables*

Variable	Definition
Dependent variable	
lnhrwage	natural logarithm of calculated hourly wages in Turkish Lira
Health variables	
sahgd	self-assessed health: 1= very good and/or good
hlthlm	functional limitation 1=psychological or physical limitation affecting daily life
nutriengh	enough nutrition: 1=protein intake in every two days
Other explanatory variables	
exp	work experience in years
exp2/100	$exp^2/100$
wed	present marital status: 1 = married
socsecprb	social Security : 1 = covered by the firm worked
perm	job status : 1 = permanent job
education	**highest degree completed**
techigh	last school completed: 1 = technical school
univ	last school completed: 1 = university
occupational class	**occupation in the current job**
prof	occupation: 1 = professional occupation
manag	occupation: 1 = manager
ascprof	occupation: 1 = associate professionals
craft	occupation: 1 = craft and related workers
clerk	occupation: 1 = clerks /shop and market sales workers
machop	machine operators
unskll	occupation: 1 = unskilled workers
nutriunskll	nutriengh*unskilled
industry	**Industry in which the individual is employed**
sector3	industry: 1 = manufacturing
sector5	industry: 1 = construction
sector6	industry: 1 = wholesale and retail sail/trade/ motor vehicle repair
size of employer	**number of regular paid employees in current job**
firmsize10	firm size: 1 = 1-10 employees
firmsize1119	firm size: 1 =11-19 employees
firmsize2049	firm size: 1 =20-49 employees
firmsize50p	firm size: 1 =50 or more employees

Table 2: Means of Variables

	Males	Females
Observations	**11051**	**3365**
Average Individuals per year	**5176**	**1682**
Average Year	**2,14**	**2,00**
hrwage	4,29	4,51
lnhrwage	1,15	1,15
sahgd	0,76	0,72
hlthlm	0,19	0,22
nutriengh	0,40	0,46
exp	16,54	9,20
$exp^2/100$	3,98	1,64
wed	0,68	0,71
techigh	0,13	0,08
univ	0,13	0,11
socsecprb	0,44	0,14
manag	0,05	0,03
prof	0,09	0,22
unskll	0,18	0,19
ascprof	0,08	0,13
clerk	0,06	0,16
servshop	0,15	0,15
craft	0,22	0,06
machop	0,17	0,06
nutriunskll	0,04	0,05
sector3	0,27	0,24
sector5	0,13	0,01
sector6	0,15	0,13
firmsize10	0,41	0,33
firmsize1119	0,13	0,12
firmsize2049	0,14	0,19
firmsize50p	0,31	0,35
perm	0,80	0,84

Table 3: Males (Total Sample)

	OLS	RE	FE
sahgd	0.0463**	0.0375**	0.0171
	(0.0190)	(0.0156)	(0.0169)
hlthlm	-0.0500***	-0.0236	-0.00116
	(0.0194)	(0.0153)	(0.0161)
nutriengh	0.229***	0.117***	0.0397***
	(0.0145)	(0.0114)	(0.0125)
exp	0.0463***	0.0528***	0.0567***
	(0.00291)	(0.00306)	(0.00664)
exp2	-0.0895***	-0.101***	-0.0856***
	(0.00772)	(0.00804)	(0.0159)
wed	0.172***	0.169***	0.200***
	(0.0207)	(0.0209)	(0.0396)
techigh	0.149***	0.182***	0.149
	(0.0227)	(0.0245)	(0.0981)
univ	0.344***	0.455***	0.207**
	(0.0331)	(0.0351)	(0.0940)
socsecprb	0.326***	0.251***	0.0722**
	(0.0217)	(0.0225)	(0.0338)
prof	0.608***	0.591***	0.308***
	(0.0421)	(0.0456)	(0.115)
manag	0.583***	0.538***	0.276***
	(0.0459)	(0.0470)	(0.0895)
ascprof	0.327***	0.320***	0.177***
	(0.0326)	(0.0351)	(0.0675)
craft	0.104***	0.0757**	0.0762
	(0.0272)	(0.0302)	(0.0649)
clerk	0.170***	0.174***	0.102
	(0.0371)	(0.0385)	(0.0781)
machop	0.0237	0.0355	0.0534
	(0.0275)	(0.0311)	(0.0706)
unskll	-0.0925***	-0.0957***	-0.0116
	(0.0293)	(0.0311)	(0.0617)
nutriunskll	-0.0362	-0.0441	-0.0441
	(0.0380)	(0.0318)	(0.0360)
sector3	-0.0478**	-0.0116	0.0665
	(0.0199)	(0.0230)	(0.0497)
sector5	-0.0562**	-0.0629**	0.0501
	(0.0279)	(0.0318)	(0.0674)
sector6	-0.135***	-0.120***	-0.0713
	(0.0232)	(0.0259)	(0.0555)
firmsize10	-0.112	-0.128	-0.218
	(0.0850)	(0.101)	(0.199)
firmsize1119	-0.0507	-0.0915	-0.231
	(0.0860)	(0.102)	(0.201)
firmsize2049	0.0508	0.00905	-0.189
	(0.0854)	(0.102)	(0.203)
firmsize50p	0.184**	0.125	-0.128
	(0.0847)	(0.102)	(0.205)
perm	0.222***	0.183***	0.00258
	(0.0260)	(0.0282)	(0.0465)
			$\chi^2_{25} = 314.464$ $Prob > \chi^2_{25} = 0.000$

Robust standard errors are shown in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Females (Total Sample)

	OLS	RE	FE
sahgd	-0.0361	0.0163	0.0698***
	(0.0409)	(0.0269)	(0.0266)
hlthlm	-0.0279	0.00684	0.0309
	(0.0441)	(0.0357)	(0.0383)
nutriengh	0.199***	0.0816***	0.0131
	(0.0324)	(0.0272)	(0.0306)
exp	0.0572***	0.0739***	0.119***
	(0.00658)	(0.00639)	(0.0161)
exp2	-0.137***	-0.177***	-0.276***
	(0.0239)	(0.0205)	(0.0443)
wed	0.107***	0.104***	0.129*
	(0.0323)	(0.0335)	(0.0682)
techigh	0.103**	0.191***	0.514**
	(0.0441)	(0.0496)	(0.261)
univ	0.371***	0.473***	0.606**
	(0.0519)	(0.0539)	(0.260)
socsecprb	0.387***	0.386***	0.309***
	(0.0486)	(0.0488)	(0.0818)
prof	0.506***	0.558***	0.427***
	(0.0695)	(0.0686)	(0.136)
manag	0.534***	0.523***	0.403***
	(0.0944)	(0.0824)	(0.133)
ascprof	0.276***	0.281***	0.171
	(0.0637)	(0.0674)	(0.136)
craft	-0.269***	-0.155	0.232
	(0.0883)	(0.0959)	(0.181)
clerk	0.0160	0.0864	0.199
	(0.0561)	(0.0576)	(0.125)
machop	0.140*	0.185**	0.363**
	(0.0719)	(0.0880)	(0.185)
unskll	-0.0292	-0.0613	-0.0316
	(0.0697)	(0.0653)	(0.115)
nutriunskll	-0.110	-0.00942	0.0365
	(0.0826)	(0.0723)	(0.0856)
sector3	-0.192***	-0.163***	0.102
	(0.0475)	(0.0550)	(0.133)
sector5	0.0812	0.0318	0.0139
	(0.145)	(0.126)	(0.137)
sector6	-0.117**	-0.0173	0.152
	(0.0496)	(0.0528)	(0.125)
firmsize10	0.00312	-0.0114	0.179
	(0.154)	(0.201)	(0.326)
firmsize1119	-0.0177	-0.0416	0.177
	(0.157)	(0.202)	(0.327)
firmsize2049	0.161	0.0876	0.172
	(0.156)	(0.201)	(0.329)
firmsize50p	0.219	0.182	0.264
	(0.155)	(0.196)	(0.308)
perm	0.495***	0.372***	-0.00830
	(0.0622)	(0.0643)	(0.128)
			$\chi^2_{25} = 108.874$ $Prob > \chi^2_{25} = 0.000$

Robust standard errors are shown in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7. References

Amemiya, Takeshi, and Thomas E. MaCurdy. 1986. Instrumental Variable Estimation of an error-components model. *Econometrica,* 54: 869-880.
Becker, Gary. S. 1962. Investment in human capital: A theoretical analysis.*Journal of political economy*, *70*(5, Part 2), 9-49
Breusch, Trevor S., Grayham E. Mizon, and Peter Schmidt. 1989. Efficient estimation using panel data. *Econometrica* 57: 695-700.
Bound, John. 1991. Self-Reported Versus Objective Measures of Health in Retirement Models. *Journal of Human Resources,* 106-138.
Cai, Lixin. 2007a. Effect of Health and Wages of Australian Men. *Melbourne Institute Working Paper Series* (2).
Cai, Lixin. 2007b. The Relationship between Health and Labour Force Participation: Evidence from Panel Data Simultaneous Equation Model. *Melbourne Institute Working Paper Series* (1).
Contoyannis, Paul, and Nigel Rice. 2001. The Impact of Health on Wages: Evidence from the British Panel Survey. *Empirical Economics*, 26: 599-622.
Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications.* Cambridge University Press: Cambridge.
Currie, Janet, and Brigitte C. Madrian. 1999. Health, Health Insurance and the Labor Market. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3C, pp. 3309-2407). North Holland: Amsterdam.
Dasgupta, Partha. 1997. Nutritional Status, the capacity for work and poverty traps. *Journal of Econometrics,* 77(1): 5-37.
Grossman, Michael. 1972. *The demand for health: A theoretical and empirical investigation.* Columbia University Press for the National Bureau of Economic Research: New York.
Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271.
Hausman, Jerry A., and William E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica,* 49: 1377-1398.
Heckman, James J. 1978. Dummy Endogenous Variable in a Simultaneous Equation System. *Econometrica,* 46: 931-939.
Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2003. Fifty years of Mincer Earnings regressions. *NBER Working Paper Series, Working Paper*: 9732.
Jäckle, Robert, and Oliver Himmler. 2010. Health and Wages: Panel data estimates considering selection and endogeneity. *Journal of Human Resources,* 45(2): 365-406.
doi:10.3368/jhr.45.2.364
Lee, Lung-Fei. 1982. Health and Wage: A Simultaneous Equation Model with Multiple Discrete Indicators. *International Economic Review* 23: 199-221.
Leibenstein, Harvey. 1957. *Economic backwardness and economic growt: Studies in the theory of ofeconomic development.* Wiley &Sons: New York.
Luft, Harold S. 1975. The Impact of Health on Earnings. *The Review of Economics and Statistics,* 57(1): 43-57.
Mincer, Jacob. 1958. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy* 66: 281-302.
Mincer, Jacob. 1974. *Schooling, Experience and Earnings.* National Bureau of Economic Research: New York.
Mushkin, Selma J. 1962. Health as an Investment. *Journal of Political Economy*, 70(5): 129-157.
Semykina, Anastasia, and Jeffrey M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics,* 157(2): 375-380.
https://doi.org/10.1016/j.jeconom.2010.03.039
Strauss, John, and Duncan Thomas. 1998. Health, nutrition and economic development. *Journal of Economic Literature* 36: 766-817.
Thomas, Duncan, and John Strauss. 1997. Health and Wages: Evidence on men and women in urban Brazil. *Journal of Econometrics,* 77: 159-185.
Thomas, Duncan, and Elizabeth Frankenberg. 2002. Health, nutrition and prosperity: a microeconomic perspective. *Bulletin of the World Health Organization,* 80(2): 106-113.
Wooldridge, Jeffrey M. 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics,* 68(1): 115-132.
Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, The MIT Press: Massachusetts.

APPENDIX
Hausman Test:

As explained by Wooldridge (2002)

Basic Random Effects Assumptions:

1. a) Strict exogeneity: $E(\upsilon_{it}|x_{it}, \alpha_i) = 0$,

   b) Orthogonality: $(\upsilon_{it}|x_{it}, \alpha_i) = 0$

2. Rank condition: $E(X'_t \Omega^{-1} X_i) = K$, where Ω is unconditional variance matrix of v_i where $\Omega \equiv E(\varepsilon_i {\varepsilon'}_i)$.

3. Efficiency : $(\varepsilon_i {\varepsilon'}_i|x_i, \alpha_i) = \sigma_u^2 I_T$ and $E\left(\alpha_i^2 \middle| x_i\right) = \sigma_\alpha^2$

Given the assumptions above and considering a case in which x_{it} includes only time varying elements, which are the only coefficients that can be estimated using the fixed effects, then:

A$var\left(\hat{\beta}_{FE}\right) = \sigma_\varepsilon^2 \left[E\left(\ddot{X}_i' \ddot{X}_i\right)\right]^{-1}/N$ and A$var\left(\hat{\beta}_{RE}\right) = \sigma_\varepsilon^2 \left[E\left(\check{X}_i' \check{X}_i\right)\right]^{-1}$ where the tth row of $\ddot{X}_i$ is $x_{it} - \bar{x}_i$ and the tth row of $\check{X}_i$ is $x_{it} - \lambda \bar{x}_i$ where $\lambda = 1/\sqrt{\eta}$ and $\eta = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + T\sigma_\alpha^2)$

Let $\hat{\delta}_{RE}$ represent the vector of random effects estimates without the coefficients for time-constant variables or aggregate time variables and let $\hat{\delta}_{FE}$ represent the corresponding fixed effects estimate; let each be $M \times 1$ vectors. Then

$$H = (\hat{\delta}_{FE} - \hat{\delta}_{RE})' \left[\mathrm{A}var\left(\hat{\delta}_{RE}\right) - \mathrm{A}var\left(\hat{\delta}_{FE}\right)\right]^{-1} (\hat{\delta}_{FE} - \hat{\delta}_{RE})$$

is distributed asymptotically as χ_M^2 under random effects assumptions.

However if the third assumption of the random effects estimation does not hold, then a robust form of the Hausman statistics is necessary; a robust Wald statistic is recommended by Wooldridge (2002) to apply this test.

Test Results:
Both the standard Hausman test and the robust Hausman test are carried out by applying user-written commands in the Stata software.

Men:
The Standard Hausman test: $\chi_{25}^2 = 480.78$, $Prob > \chi_{25}^2 = 0.000$; FE is the true model.
Robust Hausman test: $\chi_{25}^2 = 314.464$, $Prob > \chi_{25}^2 = 0.000$; this confirms that FE is the true model.

Women:
The Standard Hausman test; $\chi_{25}^2 = 147.67$, $Prob > \chi_{25}^2 = 0.000$; FE is the true model.
Robust Hausman test : $\chi_{25}^2 = 108.874$, $Prob > \chi_{25}^2 = 0.000$; this confirms that FE is the true model.