In: Secondary Education in the 21st Century
Editors: Daniel S. Beckett
ISBN: 978-1-61728-912-5

Chapter 4

SOCIAL RETURNS TO EDUCATION IN TURKEY: SOME QUANTILE ESTIMATES

Ozan Bakis*[2]*, Nurhan Davutyan**[1]*,
***Haluk Levent*[2] *and Sezgin Polat*[2]**
[1]Marmara University
[2]Galatasaray University

ABSTRACT

This paper studies local human capital externalities and returns to education in Turkey. Data comes from 2006 Household Labor Survey. Instrumental Variables-OLS estimation indicates internal (external) returns amounting to 4.9% (2.4%), while IV estimates using quantile regression range from 3% to 6.9% (1.3% to 3.5%). We discuss further characteristics of the Turkish labor market segmented by gender and show that external returns are uniformly higher for women. Our results also indicate both internal and external returns increase or equivalently the wage distribution spreads out as education increases.

1 INTRODUCTION

According to endogenous growth theory, human capital as well as research and development (R&D), are the main determinants of growth. In Romer (1986, 1990), Lucas (1988) and Aghion and Howitt (1992) knowledge is the main source of growth, but the precise mechanism leading to growth differs slightly from one work to the other. Spillovers or external effects of human capital underlie –at least if one does not take into account semantic concerns– the same thing: factor payments are higher than what is warranted by strict marginal productivity of workers. Hence, the amount of human capital or knowledge is vital for society.

In practice, education or schooling is the most important means of generating knowledge. Further, as numerous empirical works suggest, formal schooling is an important determinant of productivity levels[1]. As pointed by Mankiw *et al.* (1992) as well as Krueger and Lindahl (2001), differences in human capital are seen as the main source of income disparities. Social returns to education can be decomposed into two parts: individual or internal returns analyzed by the Mincerian approach and external returns due to spillovers, emphasized by endogenous growth theorists. Even though there is no consensus on their magnitude as yet, the existence of such knowledge spillovers is not in doubt. Otherwise, what would be the major contribution to economics of endogenous growth theory which is based on externalities?

In this paper we are interested in social returns to higher education in Turkey. Our objective is policy oriented: given the high share of the young within the general population, what should be the educational priorities of Turkish governments? Is it desirable that all young people have college degrees? Considering the necessity to deploy scarce investment funds optimally, it is essential to invest the extra Turkish Lira (TL) where its expected return is highest. This requirement implies investigating the magnitude of social returns to education carefully.

We use the 2006 Household Labor Survey. We focus on workers in the private sector where productivity considerations can be expected to dominate remuneration decisions. This enables us to analyze the impact of external returns to higher education on the wage structure in local labor markets. We also discuss the implications for each gender separately.

Ordinary Least Squares (OLS) estimates show positive and important local human capital externalities due to average level of schooling in a region. In order to estimate external returns to schooling correctly we need to estimate private or internal returns as well. According to OLS estimates an extra year of schooling is associated with 4.9% increase in private returns; similarly one percentage point increase in regional college share raises average wages by 1.3%. The corresponding quantile regression estimates extend from 2.9% to 7% for private returns and 0.76% to 1.6% for external returns[2]. If the wage distribution were normal the OLS estimates would coincide with the median quantile estimates. As explained in Section 4, the null of normality is soundly rejected in favor of a skewed and fat-tailed wage distribution. Thus the quantile median estimates (3.8% for internal and 1% for external returns) are more reliable as measures of "typical" impact.

Endogeneity is the major problem in measuring social returns to education empirically. Regional average wages may be high for at least three reasons; only one of them being knowledge spillovers. The second one is unobserved city characteristics correlated with local human capital stock that make workers more productive. Consider two regions, A and B. If an unobserved factor like amenity is higher in A and skilled workers prefer that amenity, then, skilled workers will *ceteris paribus* choose region A to work/live in. As skilled workers will be more productive, average wages in region A will be higher even without any externality. In order to get unbiased estimates of human capital spillovers, we need to control for such regional characteristics. The third one is unobserved individual characteristics e.g. ability or motivation correlated with earnings and education that make wages higher. An oft given example is Silicon Valley. The industrial structure of Silicon Valley rewards talent, leading to

1For instance, Romer (1989), Mankiw et al. (1992), Benhabib and Spiegel (1994), Barro (1999), Krueger and Lindahl (2001) and Vandenbussche et al. (2004) find that schooling is positively correlated with GDP per worker.

2 OLS results pertain to the mean of the dependent variable. Quantile regression allows a richer analysis. Here we estimated the impact of schooling and college share on the 10'th, 25'th, 50'th - median, 75'th and 90'th percentiles of the wage distribution. We provide a brief overview of quantile regressin in Section 4.

a higher than average ratio of talented workers there. More generally big cities attract the ambitious whereas less driven people may prefer a quieter provincial life. If we do not control for such self sorting, our estimates of individual and by extension social returns will be upwardly biased.

Therefore, higher average wages in a region do not necessarily reflect knowledge spillovers. There can be unobserved individual or regional variables that make earnings higher in that area. In order to control for such unobserved effects, we adopt the Instrumental Variables (IV) method. In principle we need to instrument for individual and for regional human capital levels. These instruments must be correlated with individual and regional human capital but uncorrelated with other unobserved factors that affect wages. Thus, we need an instrument for individual knowledge and another one for aggregate or regional human capital.

On the other hand finding suitably exogenous instruments is notoriously difficult in social sciences where most phenomena are interdependent. The major problem is finding an instrument affecting the dependent variable **exclusively** through the relevant endogenous variable. Fortunately, in the case of individual knowledge there exists substantial evidence that the "upward ability bias" is of about the same order of magnitude as the downward bias caused by measurement error in educational attainment. This is the conclusion reached by Krueger and Lindahl (2001, p1101) who survey the returns to education literature. As noted by the same authors, Griliches (1977), who was the first to treat the problem thoroughly, is of the same opinion. This leaves instrumenting for regional human capital. Following Vandenbussche et al (2006), we instrument aggregate or regional knowledge by its own lagged value by three years. Using such a predetermined variable is valid in the absence of serial correlation.

IV-OLS estimates yield similar results: an additional year of schooling raises wages by 4.9% points and a percentage point increase in college share is associated with 2.4% point increase in average wages. The corresponding IV median estimates are lower: 3.8% internal and 1.9% external. IV quantile estimates vary from 3% to 6.9% for internal and 1.3% to 3.5% for external returns. Thus our spillover estimates for Turkey exceed those reported by Moretti (2004b), who finds such returns to be about 1% using US data. However in view of the fact that returns to capital in emerging markets are generally higher, this discrepancy is not surprising.

Early literature on human capital spillovers relied on macroeconomic data and cross-country regressions (e.g., Romer (1989), Benhabib and Spiegel (1994) Barro (1999). These papers are especially interested in whether it is *stock* or *accumulation* of human capital that creates externalities. These studies have been criticized on a variety of grounds including data and identification problems as well as heterogeneity bias, (Krueger and Lindahl (2001), (Sianesi and Van Reenen (2003)).

Starting with Rauch's (1993) pioneering work, there is a growing literature investigating these issues using disaggregated data. Identifying knowledge spillovers using cross country data is highly problematic. Therefore, this emerging literature assesses human capital externalities occurring either within sub-national regions (Rauch (1993), Acemoglu and Angrist (2000), Moretti (2004a, b)) or within industry sectors (Sakellariou (2001), Sakellariou and Maysami (2004), Winter-Ebmer (1994), Kirby and Riley (2008)).

Rauch (1993) uses US data pertaining to Standard Metropolitan Statistical Areas (SMSA), and generalized least squares. He finds there are local external effects to average schooling, while average experience level does not matter. Rauch circumvents the endogeneity of human capital, by using a random effects model to reflect unobserved city characteristics. One major shortcoming of this approach is the assumption of zero correlation between the random effect and other explanatory variables.

Based on US data, Acemoglu and Angrist's (2000) OLS results are similar to those reported by Rauch (1993). However when they address the endogeneity problem using instrumental variables, they find no significant human capital externalities. They use state compulsory attendance laws and child labor laws as instruments of average human capital of US states, because such laws are correlated with future human capital levels and are exogenous to future adult wages. Their results are based on a sample of white men aged 40-49 from the 1960-80 Censuses. An interesting point of the paper is the use of quarter of birth as an instrument for individual schooling.

Moretti (2004b) utilizes instrumental variables to address the possible endogeneity of metropolitan college shares. His instruments are regional age structure and the presence vs. absence of land-grant colleges in American SMSAs. In the first case, the secular increase in college education leads to better educated younger cohorts throughout the USA. This gives rise to a valid instrument which is uncorrelated with other regional characteristics. In the latter case the presence of land-grant colleges –established more than 100 years ago, thus independent from current conditions– instruments for local human capital. He finds that a one percentage point increase in college share, after controlling for private returns, raises average wages by 0.6% to1.3%.

As to the relevant literature for Turkey, we note the absence of any work on external returns. Ours is the first in that regard. However there are two important papers assessing the Turkish private return to education. Tansel (2004) combines data from two sources: 1987 Household Expenditure Survey (HES) and 1989 Household Labor Force Survey (HLFS). Using HES, she estimates the private return to schooling for various education levels[3], separately for each gender. The method is joint maximum likelihood estimation of the gender and wage equations. For men, Tansel (2004) finds a rate of return between 1.9 and 16.2 % for wage earners (all ages included). The same author applies Heckman's two-step estimation method to HFLS data. Results for wage earners are very close to the ones obtained via HES data. For self-employed people, private returns range from 6.14 to 14.7 % per school year.

The second work is due to Guner and Duygan (2005) who use 2002 Household Income and Consumption Expenditures Surveys (HICES).

They restrict their sample to males aged 20 to 54. Adopting the standard Mincerian framework, they regress the logarithm of annual wage earnings[4] against years of schooling, years of experience, years of experience squared. They find that one extra year of schooling increases earnings by 12.57% – on average –.The papers by Tansel (2004) and Guner and Duygan (2005) provide careful estimates of private returns to education in Turkey. However, social returns to education were hitherto neglected for the Turkish case. In this sense our paper is a first attempt.

Finally there is a burgeoning political economy literature investigating the emergence of regional industrial centers in Turkey in the post 1980 period. The main concern of this

3Identified levels are primary, middle, high, vocational high schools and university.

4The use of total wage is problematic; in fact one needs a good measure of marketed education return; i.e. hourly wages.

literature is to explain how the liberalization measures of 1980 in conjunction with other wide ranging phenomena like the globalization process or returning expatriate workers from Germany starting businesses in their Anatolian hometowns, gave rise to the so-called Anatolian tigers. Demir et al's (2004) paper is a good example of this strand. Although not directly concerned with returns to education, there is an overlap since correct modeling of regional economic structures is a precondition for estimating social returns, i.e. internal plus external, to education accurately.

2. MODELING

Our model is adapted from Rauch (1993) and Moretti (2004a). We need it to test for the presence of external returns to human capital in different localities. This entails observing higher nominal wages in more developed regions. Then the question becomes why nominal wages are not equalized across the country. The answer being migration towards regions with high wages raises residential and commercial rents thus equalizing "amenity adjusted" real wages. As a result, in terms of utility, a worker will be indifferent between high wages/high rents and low wages/low rents.

Consider a country with N regions indexed by I, $i=1...N$. Region i has a fixed amount of land t_i and a local public good (or amenity) a_i , which households receive free of charge. There are many firms and households who can migrate at zero cost between regions. Households have different levels of human capital. Household j has a level of human capital (or equivalently, efficient labor) h_j which she supplies inelastically at wage per efficiency unit,w_i in region i.

There is a single and nationally traded consumption good y produced by capital, labor and land under perfect competition. Consumption good's price is normalized to unity, p=1. Returns to scale are constant in private inputs: labor, land and capital. The representative firm in region i has the following production function:

$$y_i=A(h_i)F(k_i,h_i,t_i)$$

$A(h_i)$ denotes the externality effect that depends on the aggregate level of human capital in region i. Individual firms do not have control on it. The important point being the rental price of capital, r, is common to all regions while the prices of land and labor are region specific. This is due to land and labor being traded locally, while capital nationally. Land's rental price is denoted as z_i in region i.

Preferences are all identical and homothetic across households. Households get utility from land, amenity and consumption good. The representative household living in region i has the following utility function:

$$U_i=u(a_i,y_i,t_i)$$

It is standard to derive cost function for firms and indirect utility function for households. Under constant returns to scale and perfect competition, unit cost is given by (we neglect r as it is common to all regions),

$p=1=C_i(w_i,z_i,A)$ and indirect utility per efficiency unit of labor by

$V_i=V_0=v(w_i,z_i,a_i)$

The spatial equilibrium is obtained when households and firms are indifferent between regions. Common nation-wide utility of one efficiency unit of labor is denoted as V_0. Then, equilibrium implies all firms having unit marginal costs and all households receiving V_0 for one efficiency unit of labor.

Consider two regions, A and B such that amenity in region A is greater than amenity in B, i.e. $a_A>a_B$. We will assume that skilled workers value amenity more than unskilled ones. Then, *ceteris paribus* a skilled worker will accept a lower wage in region A in comparison with region B. If there is no spillover effects due to local human capital, wages of skilled workers would be lower in region A. The only way we can expect higher wages for skilled workers in region A – compared to region B- are externalities due to local human capital.

Since workers are mobile between regions, in equilibrium, in every region there will be workers from each human capital level. If spillover effects increase nominal wages of skilled workers, say in region A, than rental prices should also be high in this region to ensure worker indifference between regions.

3. IDENTIFICATION

If there are human capital externalities, higher wages will prevail in more developed regions. However as already mentioned, wage differentials may also stem from some unobserved heterogeneity between regions. We can regroup these unobserved factors in two groups: individual unobserved factors or region-specific characteristics.

In order to identify human capital spillovers, we need instruments that are uncorrelated with unobserved factors raising wages, but correlated with individual and regional human capital levels. The model we use to estimate social returns is:

[1] $w_{ij} = \beta X_{ij} + \lambda R_j + \alpha H_j + \mu_{ij}$

X: holds for all individual attributes including age, tenure, social security status[5] and individual human capital i.e. years of schooling. H: holds for regional human capital proxied by college share in total employment, R: stands for regional indicators that can be correlated with regional human capital H. w: stands for logged wages.

The error term can be seen as a linear combination of three components: an unobserved individual factor θ (such as ability) that is possibly correlated with individual human capital;

5 The extent of informality in Turkey is considerable. Workers in that sector lack social protection which is reflected in their wages. See Davutyan (2008).

an unobservable regional factor μ_j possibly correlated with the level of skilled labor in that region; and an independently and identically distributed shock, ε.

[2] $u_{ij}=\mu_j+\theta_i+\varepsilon_{ij}$

Endogeneity can arise from both unobserved individual and regional characteristics. It is possible that more talented workers are more productive in a certain region A –e.g. due to technological structure complementing ability– and thus earn higher nominal wages there. However the evidence presented both by Grilliches (1977) as well as Lindahl and Krueger (2001) suggest such upward ability bias is counterbalanced by the downward bias caused by measurement error in educational attainment. Faced with the difficulty of finding a truly exogenous instrument for education, we welcome this finding.

Secondly, unobserved regional heterogeneity may bias our estimates. Some cultural, geographic etc. reason –unrelated with a region's human capital level– may account for workers being less/more productive there. Dummy variables will control for some of this unobserved heterogeneity. Regional indices can partially capture such effects. However they can not account for all omitted variables that correlate wages with average human capital in the region. Using the above notation we would write $Cov\ (\mu, H) \neq 0$.

As Acemoglu and Angrist (2000, p.2) put:

> Rauch's estimates are driven by differences in average schooling across cities. But higher incomes might cause more schooling instead of vice versa. Cities with greater average schooling may also have higher wages for a variety of other reasons. This highlights the fact that a major challenge in estimating the effects of education on income is identification.

To circumvent this source of endogeneity due to unobserved regional characteristics, we use its own lagged value as an instrument for local human capital stock[6]. This "predetermined variable" approach will yield correct estimates, provided the absence of serial correlation. We used the college share in total employment during 2003.

4. QUANTILE REGRESSION (QR)

Ordinary least squares regression focuses on the mean. Thus it only provides a partial view of the relationship between the regressors and the dependent variable. Quantile regression (QR) gives a more complete picture since it provides information about the linkage between the outcome variable (in our case wages) and the regressors at different points of the conditional wage distribution.

Other advantages of quantile or least absolute deviations regression include robustness to outliers and avoidance of distributional assumptions regarding regression errors. As pointed out by Cameron and Trivedi (2009), these features make QR especially suitable for heteroskedastic data. We performed the available tests (Breusch-Pagan / Cook-Weisberg) to

6 In an earlier version we had used employment share of women. However friendly criticsms during ERF's 13'th to 16'th July 2009 Training Workshop dissuaded us.

check for heteroskedasticity. The null hypothesis of constant variance was decisively rejected in every case. In addition both a Kolmogorov-Smirnov test as well as a skewness and kurtosis test soundly rejected the null hypothesis of normality. Consequently and noting that our median (QR) estimates of returns to education are consistently below the mean (OLS) estimates, we believe adopting this approach is appropriate as well as fruitful.

The quantile regression model posits a linear relationship between the conditional quantile and the set of explanatory variables X[7]:

[3] $w = X'\beta + \mu$

Letting F denote the (unknown) cumulative density function of w, the quantile $q \in (0, 1)$ is defined as that value of w that splits the data into the proportions q below and (1-q) above. In other words $F(w_q) = q$ and $F^{-1}(q) = w_q$. For instance since the median (q=0.5 or 50'Th percentile) of our logged wages equals 0.734, Pr (w <= 0.734) is 0.5. These concepts extend to our predictor, namely the conditional quantile regression function Q_q (w/X). Thus

[4] $Q_q(w/X = x) = x'\ \beta_q$ for $0 < q < 1$

Unlike the OLS and maximum likelihood estimators, the QR estimator does not have a closed form solution. Its computational implementation requires numerical optimization via linear programming methods. The q'th regression quantile estimate βhat_q is the solution to the following minimization problem:

[5] $$Q(\beta_q) = \sum_{w>x'\beta}^{N} q \mid w - x'\beta q \mid + \sum_{w \le x'\beta}^{N} (1-q) \mid w - x'\beta q \mid$$

When q is 0.5 equal weights are placed on both types of predictions and one gets the median estimator whose ancestry goes back to Laplace's work in 1789, Koenker (2005, p4). If, say, q = 0.75 greater weight is put on predictions where w>x'ß than for those where w<=x'ß, and we get the QR estimator for the 75'Th percentile of the wage distribution.

As Buchinsky (1998) demonstrates, the QR estimator that minimizes (5) is asymptotically normal under quiet general conditions. Thus

[6] $\beta\text{hat}_q \sim N(\beta_q, A^{-1}B\ A^{-1})$

where $A = \sum_i x_i x_i'$ $B = \sum_i f_{uq}(0/x_i)\ x_i x_i'$ and f_{uq} is the conditional density of the error term $\mu_q = w - x'\ \beta_q$ evaluated at $\mu_q = 0$. Since this term is awkward to evaluate bootstrap methods are more commonly used to generate variance-covariance estimates. This adds to the computational burden. Indeed, given that the advantages of median regression over OLS are known since the early 19'Th century, the latter's preponderance is essentially due to its having a closed form solution obviating the need for numerical optimization. From this perspective the resurgence of quantile regression is due to the microchip revolution and the concomitant cheapening of computing as well as storage capacity and the resulting emergence of large human resources data sets.

7 For convenience here we subsume regional indicators R and regional human capital H under X.

For useful overviews of the quantile method as well as additional insights, the reader is referred to Cameron and Trivedi (2005, pp89-90) and Angrist and Prischke (2009, pp269-291).

4.1. Instrumental Variables Quantile Regression

The standard quantile method, just like OLS, assumes that all the regressors are exogenous. But if as is the case here, regional human capital H is endogenous, then using conventional quantile regression to infer about its impact over the entire wage distribution will yield biased results. Chernozhukov and Hansen (2006, 2008) develop an instrumental variables quantile regression model (IVQR) that eliminates this source of bias. Given the following structural model:

[7] $w = \beta X + \alpha H + U$

Where w is logged wages, H is the endogenous human capital of the relevant region and X is the vector of control variables including regional indicators and U is an error term. The corresponding quantile regression function becomes:

[8] $Q_q (w/X, H) = \beta_q X + \alpha_q H$ for $0 < q < 1$

Chernozhukov and Hansen (2008, pp381-383) derive an estimation equation of the following form:

[9] $P (w <= \beta_q X + \alpha_q H / X, Z) = q$

using the following assumptions:

a. $\beta_q X + \alpha_q H$ is strictly increasing in q
b. U is uniformly distributed with mean 0 and SD equaling 1
c. The instrument(s) Z is independent of U
d. Z and H are dependent.

Essentially (9) provides a moment restriction which can be used to obtain IVQR estimates β_q and α_q. Specifically, for a given value of α_q. , we regress ($w - \alpha_q H$) against X and Z to estimate $\beta\text{hat}(\alpha_q)$ and $\gamma\text{hat}(\alpha_q)$ where γhat represents the estimated coefficient(s) of the instrument(s). The moment equation (9) implies that zero is the quantile solution to ($w - \beta_q X + \alpha_q H$) conditional on X and Z. Therefore that particular α_q value which makes the coefficients of the instrumental variables $\gamma\text{hat}(\alpha_q)$ as close to zero as possible, is our IVQR estimate of the impact of H on w. Formally:

[10] $\alpha\text{hat}_q = \text{arginf } W_n[(\alpha_q)] = n[\gamma\text{hat}(\alpha_q)' A(\alpha)\, \gamma\text{hat}(\alpha_q)] \quad \alpha \in \Omega$

Ω is the parameter space for α. As shown in Chernozhukov and Hansen (2008, p383), $A(\alpha)$ is the inverse of the asymptotic covariance matrix of sqrt(n)[γhat(α_q) - $\gamma(\alpha_q)$] which means that

$W_n[(\alpha_q)]$ is the Wald statistic for testing $\gamma(\alpha_q) = 0$.

In summary for a given probability level q, the estimation procedure works as follows:

a) Define a set of suitable values $\{\alpha_j, j= 1 \text{ to } J\}$ and run ordinary quantile regressions of $(w - \alpha_j H)$ against X and Z to estimate βhat(α_j) and γhat(α_j)
b) Use the inverse of the covariance matrix of γhat(α_j) to obtain the Wald statistics – namely z or F values - $W_n[(\alpha_j)]$. Take the α_j value that minimizes the $W_n[(\alpha_j)]$ as the estimate of α for that quantile level. The estimates of β_q vector are the corresponding coefficients on X.

Chernozhukov and Hansen (2008) provide two detailed examples of how their method works in practice. They graciously make their data and a Matlab program to implement IVQR available publically. However, the present study was implemented on STATA 10. Convergence was achieved without much difficulty in every case. However the estimation of the variance-covariance matrix required the use of bootstrap methods. Other studies using this same approach include Eren (2009), Galvao and Rojas (2009).

5. Regression Results

Our data set pertains to 2006 and consists of 54,728 observations on individuals working for private sector firms, in the 26 statistical regions comprising Turkey[8]. We restricted our sample to private sector employees on grounds that – as compared to the public sector- productivity considerations play a more decisive role in remuneration decisions.

We regressed logged hourly wages (WAGE) against a set of control variables. They are: the age of the worker and its squared value (AGE, AGE2), length of service in current job and its square (TENURE, TENURE2), a dummy indicating social security status (SOSEC) and a second dummy WORKPLACE to denote large enterprises with more than 50 employees. We also have data on the regional unemployment rate. (UN_RATE) for each educational stratum, a regional amenities index (INDEX), a regional openness to foreign trade index (OPEN). OPEN is calculated as per capita exports plus imports at the regional level.

We quantify each worker's individual capital stock by her/his years of schooling (EDU). Thus in each regression EDU's coefficient is our measure of private or internal return to education. The share of college graduates in each region's workforce (COLLEGE) is our proxy for the regional human capital stock. Therefore, COLLEGE's coefficient measures the external return to education.

8 The source is TURKSTAT's 2006 Household Labor Survey. It is available for a nominal fee. In EU's statistical terminology, Turkey is divided into 26 NUTS2 regions. NUTS stands for Nomenclature of Territorial Units for Statistics

Appendices 1-4 display the full set of results. It can be seen that abstracting from the amenity INDEX – discussed subsequently –, all variables have the expected signs. AGE and TENURE are strongly positively, and their squares negatively, significant. The SOSEC and WORKPLACE dummies indicate that workers with social security coverage working in large enterprises (> 50 employees) receive higher wages. The negative and significant UN_RATE (unemployment rate) coefficients point out to Phillips curve effects. The coefficients of OPEN - measuring openness to foreign trade for each NUTS region- are strongly positive. We will argue this may partially reflect an externality where the knowledge spillover is from the foreign to the domestic sector.

In what follows we highlight our salient findings and present the relevant private (EDU) and external (COLLEGE) returns to education. As mentioned previously our Kolmogorov-Smirnov and skewness-kurtosis tests strongly rejected the null of normality in favor of a skewed and fat-tailed wage distribution. The relevant p values were less than 0.0001 and 0.00001 respectively. It is well known that for such data median (rather than the mean) is the better measure of central tendency. Table 1 compares the OLS estimates with the corresponding QR or median (Q=0.5) estimates. In each case, the "more typical" median estimates of educational returns are less than the corresponding mean estimates. We believe this pattern is consistent with the leptokurtic nature of our wage distribution and thus justifies the use of quantile methods.

Table 1: Mean vs Median (Q=.05) Estimates for Social Returns. (p value <0.001 in every case)

	ALL	FEMALE	MALE
EDU ols	0.049	0.049	0.048
EDU qr	0.038	0.038	0.037
COLLEGE ols	0.013	0.023	0.009
COLLEGE qr	0.009	0.020	0.007
EDU iv-ols	0.049	0.049	0.048
EDU iv-qr	0.038	0.038	0.037
COLLEGE iv-ols	0.024	0.035	0.019
COLLEGE iv-qr	0.019	0.034	0.014
Mean logged wages (WAGE)	0.7534	0.6816	0.7730
Median logged wages (WAGE)	0.7340	0.6827	0.7340
# of workers	54,728	11,776	42,952

In view of these results we conclude that the "typical" private return to education does not vary by gender and is about 3.8%, as indicated by the EDU iv-qr row of Table 1. On the other hand, the "typical" external return to education for females is much larger than for males: 3.4% versus 1.4%. See the corresponding entries in Table 1 for the COLLEGE iv-qr row.

We now turn to anayze the returns to education information provided by the quantile estimates at the 10'th 25'th 50'th 75'th and 90'th percentiles of the wage distribution.

Table 2. Private returns to education (p value <0.001 in every case)

	Q=0.10	Q=0.25	Q=0.50	Q=0.75	Q=0.90
ALL, QR	0.0294	0.0310	0.0375	0.0517	0.0700
ALL,IVQR	0.0296	0.0312	0.0379	0.0518	0.0690
FEMALE,QR	0.0324	0.0340	0.0381	0.0515	0.0670
FEMALE,IVQR	0.0327	0.0334	0.0377	0.0511	0.0663
MALE,QR	0.0272	0.0298	0.0368	0.0519	0.0709
MALE,IVQR	0.0274	0.0297	0.0373	0.0523	0.0707

Both conventional and instrumental variables QR estimates displayed in Table 2 show that, for a given educational level (say EDU = 5 yrs), individual returns get larger as one moves up the wage distribution. This pattern holds for the whole data set as well as female and male workers separately. Moreover F-tests for the equality of these coefficients were soundly rejected. In each case the relevant p value was below 0.00001. This finding implies that as education levels increase the corresponding wage curve gets wider[9]. For instance the estimates for ALL indicate an additional year of schooling raises the lower decile of wages by 3%, the median by 3.8% but the upper decile by almost 7%. In other words wage inequality increases with education. The same finding is reported to hold for US wages starting with rhe 1990'es, Angrist and Pischke (2008, p28).

The underlying reason is still debated, however the weight of the evidence points to skill biased technical change associated with the computer revolution, Lemieux (2008, pp21-22). Given Turkey's position in and integration with the world economy (its foreign trade grew from 31% of GNP in 2000 to 45% in 2008[10]), it is tempting to argue the same phenomenon underlies our finding as well. The policy implication that follows involves promoting skills that complement and facilitate computer usage.

Table 3 lists our estimates for external returns to education. First we note our instrumental variable estimates are always higher than the corresponding conventional QR estimates. Second, as in the case of internal returns, coefficients get larger when moving up the wage curve. An F test of equality of these coefficients was decisively rejected with a p value < 0.0001 for all workers and women but not for men. Finally external returns for women are substantially above those for men. This pattern holds for each quantile level. For instance the median IVQR external returns for women is more than twice larger than the corresponding return for men – 3.4% versus 1.4%.

Table 3: External returns to education (p value <0.001 in every case)

	Q=0.10	Q=0.25	Q=0.50	Q=0.75	Q=0.90
ALL, QR	0.0076	0.0083	0.0093	0.0123	0.0164
ALL,IVQR	0.0130	0.0160	0.0190	0.0260	0.0350
FEMALE,QR	0.0061	0.0122	0.0203	0.0291	0.0392
FEMALE,IVQR	0.0100	0.0190	0.0340	0.0450	0.0620
MALE,QR	0.0050	0.0061	0.0067	0.0088	0.0099
MALE,IVQR	0.0100	0.0130	0.0140	0.0210	0.0300

9Note: In each case the quantile coefficients successively increase. Thus the "simple regression" line of WAGE against EDU for q=0.90 is steeper than the corresponding q=0.10 regression line. This means a more spread out wage distribution for higher EDU values.

10 Undersecretariat for Foreign Trade http://www.dtm.gov.tr/dtmweb/

We can think of two complementary explanations for this phenomenon. Women being relative newcomers to the labor market, have a greater zeal to learn, perhaps as a compensatory mechanism. Second, men display a greater willingness to engage in informationally meaningful interactions with female –as compared with male- colleagues. Enterprise level data could shed more light on this point. However the policy implication of this finding is clear. Social i.e. private plus external returns to education are higher for women. This finding points out to an additional reason towards promoting female labor force participation.

Two additional findings displayed in Table 4 deserve comment. These numbers are the 2SLS and instrumental variable QR estimates for OPEN and the amenities INDEX.

Table 4. The impact of trade OPENness and amenities INDEX

	2SLS	**Q=0.10**	**Q=0.25**	**Q=0.50**	**Q=0.75**	**Q=0.90**
OPEN	8.54e-06 ***	8.4e-06 ***	7.5e-06 ***	7.0e-06 ***	6.0e-06 ***	4.6e-06 ***
N= 54,728						
INDEX	-0.067**	*.01701*	*.02891*	*.01625*	-.03431*	-.15725***
N=54,728						
Underlined coefficients are INSIGNIFICANT. *** (**, *) indicates p value <= 0.001 (0.01,0.05).						

As mentioned previously, Turkey is well integrated into the world economy. The GNP share of foreign trade in 2008 was about 45%. Given the nature of the production process and of the world supply chain under globalization, a good deal of this trade is of the intra-firm or intra-industry variety. A careful study by Cakmak (2006) demonstrates that such trade involving Turkey and four major EU[11] countries increased by almost 300% from 1991 to 2004. For instance in 2004, 38% of all manufacturing trade between Turkey and France involved intra-industry trade, in the automotive subsector the ratio reached 47.2%. Similar figures hold for Germany and Italy as well. Since such trade requires exchange of personnel during the production process, we believe the significantly positive coefficients for OPEN - shown in the top portion of Table 4 - reflect knowledge spillover effects. Use of finer enterprise level data may further clarify this issue.

Finally we turn to the 2SLS and IVQR coefficients of our amenities INDEX. According to the theory expounded in our Section 2, "compensating wage reductions" would occur to equalize wages across different NUTS regions. Basically, the greater the "free" amenities of a region – such as good schools, cultural facilities, good beaches etc- valued by workers, the greater would be the wage cut involved, ceteris paribus. Thus we expected a negative coefficient for our amenities INDEX. The OLS estimates (Appendix 1, top panel) are significant and positive whereas the 2SLS estimate in Table 4 has the correct negative sign and is significant. Turning to the IVQR estimates, we note that only the 75'th and 90'th percentile estimates are negative and significant. We conclude that our "compensating wage reductions" explanation is valid only for the upper reaches of the wage distribution.

11 France, Germany, Italy and the UK.

6. SUMMARY

Endogenous growth theory emphasizes the role of knowledge spillovers in fostering economic growth. Given the econometric difficulties involved in quantifying such external effects using cross country data, recent work has focused on sub-national regions, e.g. Moretti (2004a,b).

Using 2006 data on the 26 Turkish NUTS regions and 54,728 observations we provide strong evidence on the existence of external returns to education. Our IV median estimate indicates a one percentage point increase in the ratio of college educated workers in a region, raises the "typical" worker's wages by 1.9% (3.4% females, 1.4% males). The comparable mean, namely IV-OLS , estimate is larger, 2.4%. However given the skewed and leptokurtic nature of our data, we believe ordinary least squares estimates will tend to exagerate the typical impact. LAD or least absolute devistions estimation is more appropriate for such data.

In addition LAD allows estimating the impact of the regressors on various quantiles of the wage distribution. This allows for a richer analysis. In fact our quantile estimates show that both internal and external returns to schooling increases with higher education levels. This implies rising wage inequality. Similar findings are reported for the US wage distribution since the 1990'es, Angrist and Pischke (2009), Lemieux (2008). Usually this phenomenon is ascribed to skill biased technical change associated with the computer revolution. The policy implication involves training and education programmes consistent with such technical change in production technology.

Finally our quantile estimates for external effects indicate knowledge spillovers involving women are twice larger than for men. Whatever the reasons, this finding provides an additional rationale towards promoting female labor force participation.

REFERENCES

Acemoglu, D. and J. Angrist (2001): "How large are human capital externalities? Evidence from compulsory schooling laws," in *NBER Macroeconomics Annual 2000,* ed. by B. Bernanke and K. Rogoff. Cambridge: MIT Press, 9-59.

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

Angrist, J. and J.S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton NJ.

Barro, R. J. (1999): "Human capital and growth in cross-country regressions," *Swedish Economic Policy Review*, 6, 237-277.

Benhabib, J. and M. Spiegel (1994): "The role of human capital in economic development: evidence from aggregate cross-country data," *Journal of Monetary Economics,* 34, 143-174.

Buchinsky, M. (1998): "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research", *Journal of Human Resources*, 33:88-126.

Cakmak, A. O (2006): "Pattern of Intra-industry Trade in Manufacturing between Turkey and Germany, Italy, France and the UK: 1991-2004". *Ekonomik ve Sosyal Arastirmalar Dergisi*, 3(1): 30-47 (in Turkish)

Cameron, A.C. and P.K. Trivedi (2005): *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge UK.

Cameron, A.C. and P.K. Trivedi (2009): *Microeconometrics using Stata*, STATA Press, Texas.

Chernozhukov, V. and C. Hansen. (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models", *Journal of Econometrics*, 132 (2): 491-525.

Chernozhukov, V. and C. Hansen. (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach", *Journal of Econometrics*, 142(1): 379-398.

Davutyan, N. (2008): "Estimating the Size of Turkey's Informal Sector: An Expenditure Based Approach", Journal of Economic PolicyReform, 11(4): 261-271.

Demir, O., M. Acar and M. Toprak (2004): "Anatolian Tigers or Islamic Capital: Prospects and Challenges", *Middle Eastern Studies*, 40(6): 166-188

Duygan, B. and N. Guner (2006): Income and Consumption Inequality in Turkey: What Role does Education Play? in *The Turkish Economy: The Real Economy, Corporate Governance and Reform and Stabilization Policy*, S. Altug and A. Filiztekin (eds.), pp. 63–91,Routledge, 2006.

Eren, O (2009): "Ability, Schooling Inputs and Earnings: Evidence from the NELS", Discussion Paper, Dept. of Economics, University of Nevada- Las Vegas.

Galvao, A.F and G.M. Rojas (2009): "Instrumental Variables Quantile Regression for Panel Data with Measurement Errors", Discussion Paper 09/06, Dept. of Economics, City University London

Griliches, Z. (1977): "Estimating the returns to schooling: some econometric problems", *Econometrica*, 45(1):1-22.

Kirby, S. and R. Riley [2008]: "The external return to education: UK evidence using repeated cross sections", *Labor Economics*, 15: 619-630.

Koenker, R. (2005): *Quantile Regression*, Cambridge University Press, Cambridge UK.

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

Lemieux, T. (2008): "The Changing Nature of Wage Inequality", *Journal of Population Economics*, 21:21-48.

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

Mankiw, N. G., D. Romer and D. N. Weil (1992): "A contribution to the empirics of economic growth," *Quarterly Journal of Economics*, 107: 407-437.

Moretti, E. (2004a): "Human capital externalities in cities," Chapter 51, Vol. 4, in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J. F. Thisse. Amsterdam: Elsevier.

Moretti, E. (2004b): "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 121: 175-212.

Rauch, J. E. (1993): "Productivity gains from geographic concentration of human capital: evidence from the cities," *Journal of Urban Economics*, 34: 380-400.

Romer, P. M. (1986): "Increasing returns and long-run growth," *Journal of Political Economy*, 94: 1002-1037.

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

Romer, P. M. (1990): "Endogenous technological change," *Journal of Political Economy*, 98: 71-102.

Sakellariou, C. (2001): "Identifying the external effects of human capital: a two-stage approach.", *Applied Economics Letters*, 8: 191–194.

Sakellariou, C. and R. Maysami. (2004): "Lucas type external effects of human capital: strong evidence using micro data".*Applied Economics Letters*, 11: 343–346.

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

Tansel, A. (2004): "*Education and Labor Market Outcomes*", World Bank and TEPAV working paper. .

Vandenbussche, J, P. Aghion and C. Meghir (2006): "Growth, distance to frontier and composition of human capital" *Journal of Economic Growth* 11: 97-127.

Winter-Ebmer, R., 1994. "Endogenous growth, human capital and industry wages". Bulletin of Economic Research 46: 289–314.

APPENDIX 1. OLS RESULTS

Dependent Variable: **WAGE** (logged wages)

	ALL	FEMALE	MALE
AGE	0.036***	0.032***	0.037***
AGE2	-0.000***	-0.000***	-0.000***
SOSEC	0.206***	0.264***	0.177***
TENURE	0.021***	0.025***	0.019***
TENURE2	-0.000***	-0.001***	-0.000***
INDEX	0.085***	0.165***	0.127***
UN_RATE	-0.015***	-0.013***	-0.017***
W_PLACE	0.165***	0.121***	0.183***
OPEN	0.000***	0.000***	0.000***
EDU	0.049***	0.049***	0.048***
COLLEGE	0.013***	0.023***	0.009***
CONS	-0.790***	-1.015***	-0.719***
N	54728	11776	42952
Adj. R2	0.357	0.409	0.347

2SLS RESULTS:

	ALL	FEMALE	MALE
AGE	0.036***	0.032***	0.037***
AGE2	-0.000***	-0.000***	-0.000***
SOSEC	0.207***	0.263***	0.179***
TENURE	0.021***	0.025***	0.019***
TENURE2	-0.000***	-0.001***	-0.000***
INDEX	-0.067**	<u>0.007</u>	<u>-0.009</u>
UN_RATE	0.019***	-0.016***	-0.020***
W_PLACE	0.167***	0.122***	0.185***
OPEN	0.000***	0.000***	0.000***
EDU	0.049***	0.049***	0.048***
COLLEGE	0.024***	0.035***	0.019***
CONS	0.808***	-1.025***	-0.737***
N	54728	11776	42952
Adj. R2	0.355	0.407	0.345

In both panels * indicates p value <= 0.001 ** indicates p value <= 0.01 * indicates p value <= 0.05. Underlined coefficients are INSIGNIFICANT**

APPENDIX 2. ORDINARY & IV QR ESTIMATES: ALL

Dependent Variable: **WAGE** (logged wages)

ORD-QR	Q=0.1	Q=0.25	Q=0.5	Q=0.75	Q=0.9
AGE	.05342	.04382	.03962	.03393	.01769
AGE2	-.00066	-.00053	-.0004	-.00032	*-.00004*
SOSEC	.38645	.24580	.16501	.11284	.06812
TENURE	.01200	.01438	.01833	.02509	.02762
TENURE2	-.00027	-.00024	-.00025	-.00036	-.00042
INDEX	.11003	.15971	.13763	.11377	*.05645*
UN_RATE	-.00700	-.01011	-.01149	-.01709	-.02249
W_PLACE	.16740	.15181	.13256	.13092	.14605
OPEN	8.9e-06	8.5e-06	8.2e-06	8.0e-06	7.3e-06
EDU	.02937	.03099	.03755	.05170	.06999
COLLEGE	.0076	.00828	.00929	.01235	.01645
CONS	-1.5147	-1.0203	-.71119	-.44568	*-.00821*
Psdo.R2	0.2404	0.2044	0.1781	0.2004	0.2312
N	54728	ALL individuals			

IV-QR	Q=0.1	Q=0.25	Q=0.5	Q=0.75	Q=0.9
AGE	.05262	.04334	.03896	.03345	.01787
AGE2	-.00064	-.00051	-.00043	-.00031	*-.00004*
SOSEC	.38546	.24720	.16845	.11534	.07360
TENURE	.01231	.01422	.01826	.02471	.02823
TENURE2	-.00028	-.00023	-.00024	-.00034	-.00043
INDEX	*.01701*	*.02891*	*.01625*	-.03431*	-.15725
UN_RATE	-.00886	-.01261	-.01434	-.02030	-.02654
W_PLACE	.16726	.15456	.13320	.13538	.14784
OPEN	8.4e-06	7.5e-06	7.0e-06	6.0e-06	4.6e-06
EDU	.02959	.03123	.03791	.05183	.06900
COLLEGE	.01300	.01600	.01900	.02600	.03500
CONS	-1.5066	-1.02395	-.72817	-.48814	.07823*
Psdo.R2	0.2236	0.1804	0.1553	0.1746	0.2063
N	54728	ALL individuals			

*In both panels underlined coefficients are INSIGNIFICANT. * (**) indicates p value <= 0.05 (0.01). For all others p value <= 0.001*

APPENDIX 3. ORDINARY & IV QR ESTIMATES: FEMALE

Dependent Variable: **WAGE** (logged wages)

ORD-QR	Q=0.1	Q=0.25	Q=0.5	Q=0.75	Q=0.9
AGE	.03014	.03719	.03637	.03312	.03176
AGE2	-.0003	-.00045	-.00043	-.00037	-.00024*
SOSEC	.56333	.36109	.22777	.12497	*.01427*
TENURE	.01811	.01969	.02236	.02795	.04215
TENURE2	-.00099	-.00079	-.0006	-.00057	-.00090
INDEX	.40938	.31019	.16695	.09892*	*-.03463*
UN_RATE	*-.00045*	-.00825	-.01083	-.01726	-.02328
W_PLACE	.1422	.12104	.0969	.07407	.05953
OPEN	.00001	9.2e-06	7.8e-06	6.7e-06	6.7e-06
EDU	.03243	.03402	.03817	.05155	.06705
COLLEGE	.00611**	.01224	.02032	.02906	.03925
CONS	-1.5380	-1.2028	-.89033	-.65931	-.49409
Psdo.R2	0.3011	0.2659	0.2150	0.2328	0.2682
N	**11776**	Only FEMALES			

IV-QR	Q=0.1	Q=0.25	Q=0.5	Q=0.75	Q=0.9
AGE	.02866	.03601	.03605	.03183	.02917
AGE2	-.00032	-.00043	-.00042	-.00033	-.00020**
SOSEC	.56283	.36278	.22475	.13102	*.02538*
TENURE	.01757	.02032	.02126	.02766	.04165
TENURE2	-.00098	-.00082	-.00059	-.00059	-.00089
INDEX	.31760	.20225	*.03888*	*-.05627*	-.22813
UN_RATE	*-.00261*	-.01047	-.01357	-.02059	-.02804
W_PLACE	.14394	.12315	.09705	.07719	.06627
OPEN	9.8e-06	8.3e-06	5.7e-06	4.4e-06	3.1e-06
EDU	.03270	.03342	.03774	.05114	.06627
COLLEGE	.01000	.01900	.03400	.04500	.06200
CONS	-1.4967	-1.1800	-.93237	-.70134	-.55074
Psdo.R2	0.2862	0.2331	0.1729	0.1761	0.2066
N	**11776**	Only FEMALES			

*In both panels underlined coefficients are INSIGNIFICANT. * (**) indicates p value <= 0.05 (0.01). For all others p value <= 0.001*

APPENDIX 4. ORDINARY & IV QR ESTIMATES: MALE

Dependent Variable: **WAGE** (logged wages)

ORD-QR	Q=0.1	Q=0.25	Q=0.5	Q=0.75	Q=0.9
AGE	.06082	.04642	.03949	.03179	.01464
AGE2	-.00075	-.00055	-.00044	-.00029	*-4.9e-06*
SOSEC	.29730	.20015	.13993	.10757	.07252
TENURE	.01013	.01226	.01716	.02299	.02468
TENURE2	-.00017*	-.00016	-.00019	-.00030	-.00034
INDEX	.14899	.19151	.17453	.14898	.13148**
UN_RATE	-.0100	-.01111	-.01280	-.01793	-.02198
W_PLACE	.18841	.16618	.15145	.15160	.17289
OPEN	8.7e-06	8.4e-06	8.3e-06	8.0e-06	7.3e-06
EDU	.02721	.02976	.03684	.05190	.0709
COLLEGE	.00499	.00606	.00666	.00885	.00987
CONS	-1.4959	-.98513	-.64189	-.35300	.10556*
Psdo.R2	0.2265	0.1916	0.1721	0.1939	0.2257
N	42952	Only MALES			

IV-QR	Q=0.1	Q=0.25	Q=0.5	Q=0.75	Q=0.9
AGE	.06028	.04656	.03972	.03128	.01396
AGE2	-.00075	-.00055	-.00044	-.00028	*6.6e-06*
SOSEC	.29575	.20238	.14212	.10905	.07953
TENURE	.01017	.01189	.01710	.02265	.02471
TENURE2	-.00017**	-.00015	-.00019	-.00029	-.00035
INDEX	.06540*	.07417**	.08210	*.00538*	-.09022**
UN_RATE	-.01181	-.01363	-.01510	-.02091	-.02640
W_PLACE	.18824	.16895	.15200	.15487	.17464
OPEN	8.3e-06	7.7e-06	7.4e-06	6.5e-06	4.6e-06
EDU	.02742	.02973	.03728	.05234	.07073
COLLEGE	.01000	.01300	.01400	.02100	.0300
CONS	-1.4881	-.99166	-.66777	-.38482	*.03266*
Psdo.R2	0.2132	0.1725	0.1551	0.1743	0.2070
N	42952	Only MALES			

*In both panels underlined coefficients are INSIGNIFICANT. * (**) indicates p value <= 0.05 (0.01). For all others p value <= 0.001*