University of Wisconsin-Madison
Department of Agricultural & Applied Economics

November 2003 **Staff Paper No. 465**

Returns to Schooling, Institutions and Heterogeneous Diploma Effects: An Expanded Mincerian Framework Applied to Mexico

By

Aashish Mehta and Hector J. Villarreal

AGRICULTURAL & APPLIED ECONOMICS

STAFF PAPER SERIES

Returns to Schooling, Institutions and Heterogeneous Diploma Effects: An Expanded Mincerian Framework Applied to Mexico.

Aashish Mehta
&
Hector J. Villarreal

November, 2003.

University of Wisconsin – Madison

Abstract:

We hypothesize two sources for sheepskin effects – signaling, and diplomas tied to jobs with downwards rigid wages. These theories have implications for diploma effects not only in the first, but also the second moments of the Mincerian earnings distribution that we are able to identify using a flexible econometric specification. Idiosyncrasies in Mexican labor market and educational institutions offer a natural experiment on which to train this methodology and test these theories. Correcting for heterogeneity in diplomas, we find no evidence of sheepskin effects, except on graduation from primary school. We find compelling evidence that returns to education (in both moments) are linked with labor market institutions and job-specific diplomas in the manner we hypothesize. Our econometric structure corrects for sample selectivity due to unemployment and allows us to observe behavior on the quantity axis of a labor market segmented by sheepskin effects. We also analyze the covariates of hours worked which helps to explain observed patterns in hourly earnings.

Acknowledgements:

We would like to thank Jean-Paul Chavas, Ian Coxhead, Michael Carter, Brian Gould and participants at our departmental seminar for their comments.

I. Introduction:

A large literature has grown around the methods of Mincer (1974), which involve regression of the natural logarithms of workers' wages against vectors containing measures of educational attainment and work experience. The coefficients on these vectors may be interpreted to yield a measure of the percentage increase in wages statistically attributable to particular elements of educational and work experience. Further, to the extent that forgone wages constitute the bulk of the indirect cost of education, predicted wage increases can be used to compute partial private rates of return to investments in education.

A sheepskin effect has traditionally been defined as a statistical tendency for the expected wage to increase significantly at diploma years as compared with non-diploma years. We generalize that definition. We define a sheepskin effect as a tendency for the first *or second* moments of the wage distribution to change significantly at diploma years as compared with non-diploma years. This is important because, as we argue below, the circumstances that can give rise to sheepskin effects in first moments have even stronger testable implications for the second moments of the wage distribution. We utilize a maximum likelihood structure to identify, estimate and test for sheepskin effects in both moments of the earnings distribution.

The Mincerian equation looks only at the price of labor, not the quantity. Thus, while intuitively finding a sheepskin effect may indicate something other than a Walrasian equilibrium in the labor market, it doesn't suffice to characterize that equilibrium. We therefore include a Mincerian specification of an employment equation. This allows us to investigate how the quantity side of the labor market adjusts to discontinuities in price.

The traditional literature on sheepskin effects is closely connected to signaling theories originating with Arrow (1973) and Spence (1974). While most empirical papers allude to the old wisdom that signaling is neither necessary (Chiswick - 1973, Hungerford and Solon - 1987) nor sufficient (Riley - 1979) for sheepskin effects, they do not present any alternatives seriously. More recent work that builds from the dynamic optimization framework originating in Becker (1967) has highlighted the role of institutions and government policy on the returns to education. Card (2001) provides an excellent survey of this literature. These institutions include minimum wages, collective bargaining and mandatory primary school attendance. These institutional effects features can be embedded in a Mincerian econometric framework.

We incorporate labor market and educational institutions into a Mincerian analysis of Mexico. This enables us to shed new light on the existence of, and reasons for, a variety of anomalies in the returns to education, including sheepskin effects. We propose an alternative, institution driven explanation for sheepskin effects, arguing that a sheepskin effect may occur if particular diplomas are tied to jobs which exhibit downwards wage rigidity. A central feature of our work is a natural experiment on the relationship between job-specific diplomas and labor market institutions that provides strong evidence in favor of this explanation of some anomalies in the Mincerian returns.

Mincerian returns are useful because they shed light on a number of issues. On a macroeconomic level, they provide an indication of the efficiency of human capital markets. In a sufficiently generalized Mincerian setting, one might spot shortages of particular skills. For example, through the 1990s the Mexican government aggressively targeted an alleged shortage of technically skilled workers. We test for shortages of technicians in Mexico in 2002 and find that such shortages, if they ever existed, are not evident in 2002.

Another, longer term, macroeconomic application of the Mincerian approach is to the question of indivisible investment goods. Azariadis and Drezen (1990) propose that the presence of indivisible investment goods may be responsible for the divergent growth experiences of countries. Galor and Zeira (1993) and Ljunqvist (1993) propose that education is such an indivisible good. We find, somewhat paradoxically, that while in our sample educational attainment is clustered at diploma years, there is no evidence of indivisibility from the rates of return, other than at the primary school level.

Mincerian returns should be of interest to policy makers, especially those involved in education, labor and welfare policy, because of their implications for income distribution (Fortin and Lemieux, 1997). This is more the case, due to the enormous impact of education on income, the fact that access to education is not uniformly distributed, and the sheer volume of public money allocated to education. Our specification helps to characterize determinants of the wage distribution including heterogeneous diplomas, labor market institutions and socio-demographic characteristics.

We examine the following issues. Conceptually - we first ask whether signaling is the definitive explanation for sheepskin effects and provide an alternative one. We then go on to formulate tests for diploma effects in second moments of the earnings distribution, and in the employment equation. Empirically – we test for the presence of sheepskin effects in Mexico. In doing so, we are careful to capture as much heterogeneity in diplomas and their recipients as possible in order to reduce the risk of false negatives. We also ask whether certain institutions – unions, public sector employment practices, minimum wages, and diploma requirements for some jobs - are important in explaining Mincerian returns. Some of them appear to be, though depending on the nature of the market, institutional effects may be eroded by market forces.

In addition to these contributions we develop a technical framework that is more general than those applied to date. Because this model nests those used for most previous Mincerian studies, it allows us to test some of their assumptions. Most importantly, we test the homoskedasticity assumption made by most authors, which we demonstrate potentially biases the wages predicted by the model, not just their standard errors. Second, we test for the presence of selectivity bias, as earnings are only defined for the employed.

We also estimate the effects of education on the number of hours worked. This is important because we find compelling evidence that education is associated with a shorter work week. Thus, education probably has welfare consequences that would be missed when focusing on hourly earnings alone.

The remainder of the paper is structured as follows. Section II develops the arguments underlying both explanations of sheepskin effects. It also details a natural experiment to test whether job-specific diplomas and downwards rigid wages generate sheepskins, and motivates tests for sheepskins in the first and second moments of the earnings distribution. Section III introduces the data and provides a description of the Mexican education system. Section IV presents the econometric specification and formalizes our hypothesis tests. Section V presents and interprets the model estimates and hypothesis test results. Section VI concludes. Two appendices provide econometric derivations.

II. Signaling and Institution Driven Sheepskin Effects

Sheepskins do not imply signaling and vice versa. To date, studies in this area have tended to state this, test for sheepskins statistically, and then stop. However, more can be said from a more detailed study. In particular, the reasons for the existence of sheepskin effects can be broadly categorized in two.

First, there is the signaling argument. Essentially, this states that diploma receipt conveys positive information to employers regarding potential employees that could not be inferred in its absence. This argument for a positive signaling effect requires two presumptions. First, that the information inferable from a diploma is considerably greater than what is inferable from the successful completion of a non-diploma school-year. Second, it requires that this information is positive. It seems uncontroversial to argue that if a diploma also carries a grade which permits wage offers to potential employees in accordance with their skills, then a sheepskin effect may not become evident in the expected earnings. Weaker diplomas will carry a negative sheepskin, while stronger ones will carry a positive one. In other words, the sheepskin effect may not be observable in the mean wage, but should be observable in the wage dispersion. The literature to date has only focused on the first moment of the wage distribution. We allow sufficient flexibility in our variance matrix to capture sheepskins in the second moments.

There is a second argument for the existence of positive sheepskin effects which, to our knowledge, has not been treated in the literature to date. It is that they might result if particular diplomas are necessary for obtaining particular jobs. A general equilibrium analysis of this argument reveals a further necessary condition: the wage paid by such jobs must be downwards rigid. For, if it were not, the supply response in a competitive labor market would erode the premium on the diploma. We refer to this argument as the job-specific diploma hypothesis.

We distinguish between teachers, technical high-school graduates and conventional high-school graduates in an effort to capture this story. This is a good natural experiment for the following reasons. Conventional high school degrees are flexible in the job opportunities they offer. Technical high school degrees and teaching certification lead to specific jobs. Teachers' earnings are strongly downwards rigid due to their position as unionized public sector employees,[1] while those of technical high school graduates probably are not. Thus the job-specific diploma argument suggests that teaching degrees could carry an unusually high premium, while technical and formal high-school diplomas should reap roughly equal returns, despite the job-specificity of technical degrees.

Note also that if jobs are tied to diplomas, we are unable to distinguish between different types of diplomas, and these different diplomas lead to jobs with varying pay scales, then we would also expect to see a sudden increase in the variance of earnings at diploma years. Thus, evidence of an increase in the variance of earnings is necessary for confirmation of a signaling theory for sheepskin effects, but it will not suffice.

Diploma effects are likely to erode with experience, as the importance of diplomas gives way to that of employer recommendations and workplace performance. They may also be specific to gender, location (urban vs. rural). Thus failure to estimate group-specific returns to education may return false negatives on sheepskin tests. We therefore partition our sample according to gender, location and cohort to test for sheepskin effects more carefully.

III. The empirical environment:

The Mexican education system is a mixture of public and private institutions. The public institutions depend on the federal, state or municipal governments for funding. Even though many children attend kindergarten, it is not an official prerequisite for admission to primary school.[2] Formal education prior to college usually requires successful completion of 12 years: 6 of primary school, 3 of junior-high and 3 of high school. College typically takes five years to complete, although the duration does vary.

Parallel to the formal education track, analogous levels exist for technical education that provide a similar curriculum to the formal school system, complemented by vocational training. We limit our analysis to technical high schools, dropping the few

[1] 139 out of 173 identifiable teachers in our sample are union members.
[2] This may vary according to states or the kind of school. Many private schools do require some preprimary education.

workers with other levels of technical education from our sample. Teachers in Mexico are trained at two levels. A four year course after junior high yields a *normalista* degree. Teacher training at the college level yields a *normalista-superior* degree, though we are unable to distinguish its recipients from other college graduates.

The data source for this study is ENIGH 2002 (Encuesta Nacional de Ingreso y Gasto de los Hogares), which is a household income-expenditure survey, collected by INEGI (Instituto Nacional de Estadistica Geografia e Informatica) in 2002. Even though very similar surveys were collected for the years 1992, 1994, 1996, 1998 and 2000, we refrain from using them because some education attainment information in these surveys was recorded categorically rather than by year.

We pare down our sample using criteria that are standard in this literature. We do not include persons in our sample who are even partially self-employed, for fear of conflating returns to unreported capital ownership, with those to education.[3] Persons under 16 (the legal working age) and over 65 years old are also dropped. Those reporting hours worked but no remuneration are stricken from the sample. Voluntarily unemployed persons - those reporting zero worked hours and not searching for a job, are also excluded. Persons working zero hours that are searching for a job are included in our analysis and categorized as unemployed. Our sample of graduate degree recipients and those with incomplete teaching or technical high-school degrees was too thin for computational purposes and we were forced to drop them. We also dropped those currently enrolled in school as their labor market decisions and opportunities are markedly different from the rest of our sample.

Our dataset has two strengths. First, it allows us to differentiate between different types of diplomas. Second, it contains data on the successful completion of school years and diplomas, rather than just temporal measures of schooling. As Jaeger and Page (1996) point out this is important because imputing completion from temporal data can bias results.

[3] Smith and Metzger (1998) find, in a Mexican context, that failure to control for returns to capital biases estimates of returns to education upwards as educational attainment correlates positively with capital and earnings.

IV. The Model and Hypothesis Tests:

The Mincerian Equation

The most common Mincerian equation takes the following form:

(1) $y = \ln w = d_o + d_E E + d_{E2} E^2 + \sum_{l=p,j,h,c} \left(d_{sl} s_l + d_{Dl} D_l \right)$, where w is a person's hourly earnings, sometimes referred to as their implicit wage. E, potential experience, is the maximum length of time they could have been in the labor force given their age and education. l indexes the level of education (primary, junior-high, high-school and college). s_l measures the number of years of education level l completed, and is therefore bounded between zero and the number of school years required to complete that level. D_l indicates whether the lth diploma was received. Because of the semi-logarithmic form of the equation, the exponential growth rate in percentage terms associated with years of experience and of schooling at level l are $d_E + 2d_{E2}E$ and d_{sl} respectively. Similarly, $\exp(d_{Dl}) - 1$ is the percentage wage increase associated with receipt of diploma l over and above that conferred by completion of the final year of the degree. Typically, d_0 is permitted to vary with personal characteristics. Notice that a specification that "corrects" for such personal characteristics through d_0 still imposes constant returns to education and experience across the sample.

We generalize this specification before embedding it in a larger 3 equation likelihood structure. Specifically, we know who in our sample received teaching certification or technical high-school degrees instead of formal high-school degrees. We allow for different returns to these degrees, as well as different returns to college for formal and technical high school graduates. Thus, our earnings equation is:

(2)
$$y = \ln w = d_o + d_E E + d_{E2} E^2 + \sum_{l=p,j,h} \left(d_{sl} s_l + d_{Dl} D_l \right) + D_h \left(d_{PF}^{sc} s_c + d_{PF}^{Dc} D_c \right)$$
$$+ D_T \left(d_{DT} + d_{PT}^{sc} s_c + d_{PT}^{Dc} D_c \right) + d_{Teach} D_{Teach},$$

where D_h, D_T and D_{Teach} are mutually exclusive indicators of receipt of a formal high-school diploma, technical high-school diploma, or a *normalista*. Recall that our sample excludes those who did not complete technical high-school. Hence students with technical degrees are characterized by $D_h = s_h = 0, D_T = 1$, a total return to the degree over the three years of $\exp(d_{DT}) - 1$, and annual average rates of return to the degree of $d_{DT}/3$.

The superscripts PF and PT designate returns post-formal high school, and post-technical high school.

Our Model:

We are interested in the determinants of three variables: employment (z_i = 0 or 1), hours worked if employed (h_i) , and the logarithm of hourly earnings if employed (y_i). In order to investigate these, we specify the following structure, based, in principle, on Heckman's (1974) selection scheme. Each person observed in the cross-section is subscripted by *i*.

$$(3) \qquad \begin{bmatrix} z_i^* \\ h_i^* \\ y_i^* \end{bmatrix} = \begin{bmatrix} b_z x_{zi} \\ b_h x_{hi} \\ b_y x_{yi} \end{bmatrix} + \begin{bmatrix} e_{zi} \\ e_{hi} \\ e_{yi} \end{bmatrix};\ e_i = \begin{bmatrix} e_{zi} \\ e_{hi} \\ e_{yi} \end{bmatrix} \sim N(0, \Sigma_i);\ \Sigma_i = \begin{bmatrix} 1 & r_{1i} q_i & r_{2i} s_i \\ r_{1i} q_i & q_i^2 & r_{3i} q_i s_i \\ r_{2i} s_i & r_{3i} q_i s_i & s_i^2 \end{bmatrix};$$

$z_i = 1$ if $z_i^* = 0$ and 0 otherwise;

$h_i = h_i^*$ if $z_i^* = 0$ and is unreported otherwise;

$y_i = y_i^*$ if $z_i^* = 0$ and is undefined otherwise.

Thus, z_i^* is latent employment propensity while h_i^* and y_i^* are the latent hours and logged earnings potentials – observable only if a worker is employed. Σ_i is a positive definite variance matrix for person *i*. q_i and s_i are the standard deviations of the "unexplained" components of the hours and logged earnings potentials respectively. Each of the $?_{ki}$ is a correlation coefficient between unobservable components. Thus, for example, finding $?_{3i}$<0 would indicate that if person *i* works less hours than our model predicts, they are also likely to earn more per hour than it predicts.

The allowance for heteroskedasticity is implemented without loss of generality via the Cholesky decomposition:

$$(4a) \qquad e_i = A_i u_i;\ u_i \sim N(0, I_3);\ A_i = \begin{bmatrix} 1 & 0 & 0 \\ a_{3i} & a_{1i} & 0 \\ a_{4i} & a_{5i} & a_{2i} \end{bmatrix};\ a_{ji} = g_j^0 + g_j x_{ji}.\ j = 1,..,5.;$$

where x_{ji} are worker characteristics that may condition the variance matrix. From (3) and (4a) it follows that:

$$\text{(4b)} \quad \Sigma_i = V(e_i) = A_i A_i' = \begin{bmatrix} 1 & a_{3i} & a_{4i} \\ a_{3i} & a_{1i}^2 + a_{3i}^2 & a_{3i}a_{4i} + a_{1i}a_{5i} \\ a_{4i} & a_{3i}a_{4i} + a_{1i}a_{5i} & a_{2i}^2 + a_{4i}^2 + a_{5i}^2 \end{bmatrix}_i = \begin{bmatrix} 1 & r_{1i}q_i & r_{2i}s_i \\ r_{1i}q_i & q_i^2 & r_{3i}q_i s_i \\ r_{2i}s_i & r_{3i}q_i s_i & s_i^2 \end{bmatrix}$$

Note that the five parameters of the variance matrix are exactly identified from the Cholesky matrix. However, a given variance matrix only identifies the magnitudes of a_1, a_2, and a_5, not their signs. For the rest of this section, it is presumed that the signs of a_1 and a_2 are constant and positive. This eliminates the need for confusing caveats when proposing hypothesis tests. The assumption that a_1 and a_2 have constant signs will be verified.

The derivation of the log-likelihood function is relegated to appendix A. Appendix B derives the following expressions for the expectation and standard deviation of hourly earnings $\left(w_i^* = \exp\left(y_i^*\right)\right)$ for person *i*.

$$\text{(5a)} \quad E(w_i^*) = \exp\left(b_y x_{yi} + s_i^2/2\right),$$

$$\text{(5b)} \quad S.D.(w_i^*) = \exp\left(b_y x_{yi}\right)\sqrt{\exp\left(2s_i^2\right) - \exp\left(s_i^2\right)}$$

This means that in the presence of conditional heteroscedastity in logged earnings (i.e. ?$_2$, ?$_4$, ?$_5$, ? 0), a homoskedastic model is incapable of predicting not only the second, but also the first moment of earnings distribution, underestimating the expected earnings for persons subject to above average wage variability. It also means that tests on b_y do not suffice to test hypotheses regarding actual wages in a heteroskedastic world.

Next, we delineate the content of the main equations and the Cholesky matrix. Three criteria were used in selecting the conditioning variables. First, would their inclusion allow us to estimate parameters crucial to our hypothesis tests? Second, would their exclusion mingle returns to education for different types of people, resulting in erroneous acceptance of the null of no sheepskin effects? Third, the consistency of our parameter estimates hinges on the veracity of our distributional assumptions (3). We would therefore like to include variables whose exclusion is likely lead to misspecification.[4]

[4] It is desirable to test the distributional assumptions by means of a Breusch-Pagan test on the estimated standardized residuals ($\hat{u}_i$). Regrettably, because z_i^* is never observed, these estimates of u_i are unobtainable.

Logged wages, y_i^*, and employment propensity, z_i^* are conditioned on exactly the components of the RHS of (2), except that a few intercept shifters are added. Each equation is shifted by gender, region, and urban vs. rural location. Union membership and holding a *Normalista* condition earnings, but not employment, as there are almost no unemployed union members or teachers in our sample. For similar reasons, it was impossible to estimate separate employment effects of college for technical and high school graduates, and they are assumed to have the same coefficients on college education.[5] Additionally z_i^* is shifted by marital status and the interaction of marital status and gender. Thus:

(6a)
$$\begin{aligned} b_z x_{zi} &= b_z^0 + b_z^M D_{Male} + b_z^R D_{Rural} + b_z^U D_{Union} + b_z^{Teach} D_{Teach} \\ &+ b_z^{North} D_{North} + b_z^{South} D_{South} + b_z^C D_{Couple} + b_z^{CM} D_{Couple} D_{Male} + b_z^E E + b_z^{E2} E^2 \\ &+ \sum_{l=p,j,h} \left(b_z^{sl} s_l + b_z^{Dl} D_l \right) + D_h \left(b_{z,PF}^{sc} s_c + b_{z,PF}^{DC} D_c \right) + D_T \left(b_z^{DT} + b_{z,PT}^{sc} s_c + b_{z,PT}^{sc} D_c \right) \end{aligned}$$

(6b)
$$\begin{aligned} b_y x_{yi} &= b_y^0 + b_y^M D_{Male} + b_y^R D_{Rural} + b_y^U D_{Union} + b_y^{Teach} D_{Teach} \\ &+ b_y^{North} D_{North} + b_y^{South} D_{South} + b_y^E E + b_y^{E2} E^2 \\ &+ \sum_{l=p,j,h} \left(b_y^{sl} s_l + b_y^{Dl} D_l \right) + D_h \left(b_{y,PF}^{sc} s_c + b_{y,PF}^{Dc} D_c \right) + D_T \left(b_y^{DT} + b_{y,PT}^{sc} s_c + b_{y,PT}^{Dc} D_c \right) \end{aligned}$$

In principle, we could have conditioned hours, h_i^*, on the same Mincerian variables. However, as we do not have good reasons to propose the possibility of diploma effects in the hours equation, we include only four slopes - one for each level of schooling, experience and its square, and the same intercept shifters as are included for the employment equation:

(6C)
$$\begin{aligned} b_h x_{hi} &= b_h^0 + b_h^M D_{Male} + b_h^R D_{Rural} + b_h^U D_{Union} + b_h^{North} D_{North} + b_h^{South} D_{South} + b_h^C D_{Couple} \\ &+ b_h^{CM} D_{Couple} D_{Male} + b_h^E E + b_h^{E2} E^2 + \sum_{l=p,j,h,c} b_h^{sl} s_l + b_h^{DT} D_T + b_h^{Teach} D_{Teach} \end{aligned}$$

The greatest drawback to using a Cholesky decomposition to impose positive definiteness on a heteroskedastic variance matrix is that it becomes impossible to condition one element of the variance matrix on an exogenous variable, without

[5] Unlike the case of union members and teachers, this is not because a tiny *fraction* of them are unemployed, but because a tiny *number* of them are. 9 out of 200 technical high-school graduates with some college education are unemployed, loosely equiproportional with 105 out of 2057 of their formal high school counterparts. Thus, imposing equality of the coefficients is unlikely to be costly.

inadvertently conditioning other variance parameters on it also. Despite these constraints, it is clear from (4b) that the variables conditioning a_1 and a_2 will most strongly effect q and s respectively. Similarly, r_1, r_2 and r_3 can be conditioned through a_3, a_4 and a_5 respectively.

We thought of likely scenarios wherein an urban location, gender and membership of the teaching profession would condition all five elements of the variance matrix. We therefore conditioned each Cholesky element on these characteristics. Similarly r_1, r_2, r_3 and q are conditioned on the number of years of schooling.[6] Unionization was supposed to effect q , s and r_3 for obvious reasons. Finally, in keeping with the discussion of section II, s was conditioned on the same variables a y_i^*, through a_2, in order to capture diploma effects in second moments. Hence we have the following equations:

(7a) $$a_{1i} = g_1^0 + g_1^M D_{Male} + g_1^R D_{Rural} + g_1^U D_{Union} + g_1^T D_{Teach} + g_1^S Schooling$$

(7b) $$a_{2i} = g_2^0 + g_1^M D_{Male} + g_1^R D_{Rural} + g_1^U D_{Union} + g_2^T D_{Teach} + 1g_2^E E + g_2^{E2} E^2 + \sum_{l=p,j,h} \left(g_2^{sl} s_l + g_2^{Dl} D_l \right) + D_h \left(g_{2,PF}^{sc} s_c + g_{2,PF}^{Dc} D_c \right) + D_T \left(g_2^{DT} + g_{2,PT}^{Sc} s_c + g_{2,PT}^{Dc} D_c \right)$$

(7c) $$a_{3i} = g_3^0 + g_3^M D_{Male} + g_3^R D_{Rural} + g_3^S Schooling$$

(7d) $$a_{4i} = g_4^0 + g_4^M D_{Male} + g_4^R D_{Rural} + g_4^S Schooling$$

(7e) $$a_{5i} = g_5^0 + g_5^M D_{Male} + g_5^R D_{Rural} + g_5^U D_{Union} + g_5^T D_{Teach} + g_5^S Schooling\ .$$

Hypothesis tests:

This econometric structure allows us to test a variety of interesting hypotheses. We break these into five groups. First, this structure nests those of most Mincerian studies, and there allows us to test some of their assumptions. To the extent that these assumptions are rejected, results to prior studies are somewhat inconsistent. We test the restrictions implied by a homoskedastic model ($g_j = 0, \forall j$) by means of a likelihood ratio test (LRT) on the difference between the homoskedastic and heteroskedastic model. The resultant LRT statistic is distributed $c^2_{(33)}$. We also test for the presence of selectivity bias

[6] The key to table 5 describes how the 'schooling' variable was constructed.

because y_i^* is undefined for the unemployed. The null hypothesis of no selectivity bias is imposed through the restriction $g_4^0 = g_4 = 0$, and the LRT statistic arising out of comparison with the full model is distributed $c_{(4)}^2$.

Second, we check our priors on the signs of some of the intercept shifters. While not all of these constitute central economic hypotheses of our paper, rejection of these common sense priors could signal problems with our specification. We therefore check to see whether the coefficients of our model support the following statements, which we believe to be true in reality: *ceteris paribus* – (i) men work more (outside the home) than women $\left(b_h^M > 0\right)$; (ii) earnings are highest in the North $\left(b_y^{North} > 0\right)$ and lowest in the South $\left(b_y^{South} < 0\right)$; (iii) unionization increases earnings $\left(b_y^U > 0\right)$, reduces hours $\left(b_h^U < 0\right)$, reduces the variance of earnings $\left(g_2^U < 0\right)$ and hours $\left(g_1^U < 0\right)$; (iv) Rural dwellers earn less $\left(b_y^R < 0\right)$; (v) Married men work longer hours than bachelors $\left(b_h^C + b_h^{CM} > 0\right)$, while the reverse is true for women $\left(b_h^C < 0\right)$. While we have no prior on the issue we do check to see whether men earn more or less than women, other things equal.

Third, we enquire after the results of our natural experiment. We begin by checking the null hypothesis that the expected log wage of formal and technical high-school graduates is the same by comparing the quantity $3b_y^{sh} + b_y^{Dh}$ with b_y^{DT}. Given the presumed absence of downwards rigidity in wages accruing to either type of graduate, our job-specific diploma hypothesis suggests they should be equal. To test it more rigorously, we re-estimate the model with the equality imposed and test the restrictions via a $c_{(1)}^2$ LRT. In order to test the equality of actual (not logged) wages predicted, we impose the additional assumption that $3g_2^{sh} + g_2^{Dh} = g_2^{DT}$, and compare the results to the full model by a $c_{(2)}^2$ LRT statistic. The premium on a four year *normalista* relative to a formal (or technical) high-school degree plus a year of college, is appreciable by comparison of b_y^{Teach} with $3b_y^{sh} + b_y^{Dh} + b_{y,PF}^{sc}$, (or $b_y^{DT} + b_{y,PT}^{sc}$). We formally test the restrictions $3b_y^{sh} + b_y^{Dh} + b_{y,PF}^{sc} = b_y^{DT} + b_{y,PT}^{sc} = b_y^{Teach}$ on the log wage premium using a $c_{(2)}^2$ LRT statistic relative to the full model. We test for a teaching premium in the actual wage premium by adding the corresponding restrictions on g_2 and testing them jointly using a $c^2_{(4)}$ LRT statistic.

Fourth, we test for sheepskin effects. Because sheepskins cannot be identified for teaching degrees and technical high school diplomas, these levels are excluded from the following tests. The test in the first moments of the logged-earnings distribution for a sheepskin effect of the *l*th diploma simply corresponds to a students *t*-test of the alternative hypothesis that $b_y^{Dl} > 0$ against the null $b_y^{Dl} \leq 0$. Similarly, the alternative hypotheses of sheepskin effects in second moments are tested via $g_2^{Dl} > 0$. Last, noting the possibility that diploma effects are expressed in terms of access to employment, we test for the presence of sheepskin effects in the employment equation $\left(b_z^{Dl} > 0\right)$.

Fifth, we examine the behavior of hours. We ask whether the length of the average work week trends with education in any interesting fashion.

Each of the above exercises is conducted on the full sample as well as sub-samples drawn according to gender, urban vs. rural location and age respectively. We do so in order to check for the possibility that pooling obscures important results.

V. Results:

We estimated the model delineated by equations (3), (4), (6) and (7) on the full sample of 16,675 workers.[7] The results are presented in table 1. All references to parameter estimates in the following section are to these numbers. We analyzed key features of the model for a variety of profiles using the delta method. Figures 1,2 and 4 provide the results for an urban, non-union, married male from central Mexico with five years of labor market experience. Figure 3 is for the same profile, although results are from estimates for the male and female sub-samples. Results for the other profiles we have looked at do not differ qualitatively. We also estimated the model on the following sub-samples for reasons outlined in section II: males, females, urban dwellers and young

[7] The implicit wage is calculated as the ratio of all wage and salary income to hours worked in the last quarter. Potential experience is: age – years of schooling – 4.

workers (those no older than 30).[8] The qualitative results are exactly the same in all samples, with the exception of those relating to the hours worked by men and women.[9]

Turning to the first group of hypotheses described in section IV, we test two maintained hypotheses of previous Mincerian studies. The null hypothesis that the model is homoskedastic is soundly rejected. This is obvious from the significance of the majority of the variables conditioning the Cholesky elements. More formally, we test this in the first row of table 2 via a LRT, and unambiguously reject the null. We note, however, that allowing for heteroskedasticity did not change our qualitative results regarding the existence of sheepskin effects in log wages. The value of the heteroskedastic model lies in its ability to capture results in second moments, as well as to consistently estimate wages (not logged). As indicated in equation (5), a homoskedastic model cannot consistently predict wages in the presence of heteroskedasticity. Figure 1 shows that the variance of log wages varies with education. Figure 3 demonstrates that the variance of hours varies with education, but only for women.

Similarly, the null that selection into employment does not bias results is rejected. Figure 4 demonstrates that for our reference profile r_2 is definitely negative. This indicates that workers with higher earnings potential are less likely to be employed, *ceteris paribus*, and that models that fail to correct for selection by employment are therefore likely to underestimate log wages. We test the null rigorously in the second row of table 2, and reject it firmly.

Turning next to the simple hypotheses regarding the mean and variance intercept shifters, we find statistically significant confirmation of each of them. From the 'logged earnings' column in table 1, we note that hourly earnings in the North are 13% higher than those in central Mexico, while workers in the South own about 20% less.[10] Union members earn an extra 43% per hour, while rural dwellers earn 34% less than their urban counterparts in nominal terms. The size of these effects gave us pause and motivated estimation of the model on sub-samples, although the qualitative effects we are interested

[8] The sample sizes for these cross sections were: 11248, 5427, 13128 and 7365 respectively.
[9] Results for sub-samples are available on request, as are post-estimation results for other profiles.
[10] For any log-wage intercept shifter, b, the percentage increase in earnings is $e^b - 1$.

in did not vary with sub-sample. From the 'hours' column, we find that on average men work 2 hours more per week than women, and married men work 4 hours more than bachelors while married women work 4.6 hours less than spinsters. Union members work 3.4 hours less per week.

The Cholesky terms a_1 and a_2 which condition the variances of hours and hourly earnings respectively are positive in every profile we looked at. It follows that the sign of an intercept shifter in these equations applies also to its impact on the corresponding variance term, although its magnitude does not. Hence, as predicted, union membership reduces the variance of earnings and hours. The confirmation of each of our priors by our model suggests that our data and specification provide a valid description of Mexican labor market outcomes.

Although we did not have a prior on this, we find that men earn around 20% more than women on an hourly basis. While this could be due to discrimination, we stress the importance of other, possibly complimentary interpretations. For example, as we will argue below, figure 3 *might* indicate that women choose different occupations, substituting flexibility in work hours for pay.

Our third set of hypotheses relate to our natural experiment. Beginning with the high-school coefficients of the 'logged earnings' equation, we find that the total Mincerian return over three years of formal high-school is 38.9%[11], compared with 38.2% for a technical high-school degree. The standard errors on the relevant coefficients suggest that this is not likely to be a statistically significant difference. In the third row of table 2 we report a p-value of 0.88 from a $c^{2}_{(1)}$ test on the null hypothesis that the returns are equal. As (5) suggests, though, a test in the predicted log wage could obscure actual wage differences only visible in the variance of log wages. A $c^{2}_{(2)}$ LRT on the null in actual wages (Table 2, row 4) finds little evidence of a disparity, with a p-value of 0.30. Note that a cursory check of the corresponding coefficients in the 'a2' column indicates that the variance of logged earnings is a little higher for technical graduates. This means that technical high-school graduates earn slightly more than formal graduates, as is visible in figure 2. This is an example of the insufficiency of the

[11] This is: $\exp\left(3b_y^{sh} + b_y^{Dh}\right) - 1$.

homoskedastic Mincerian model for testing hypotheses about the wage distribution. However the difference is not statistically significant in this case.

Figure 2 provides visual confirmation of the above results. Note also that though the difference between wages for technical and formal graduates was insignificant at the end of high-school, and annual rates of return were roughly equal at around 15%, the heteroskedastic Mincerian specification allows for a large wage difference between them by the end of college.

Turning next to those with *normalistas*, we estimate a Mincerian rate of return over 4 years of 138%. Formal and technical high-school graduates with a year of college both earn 4-year returns of 62.08% and 62.13% respectively. For rigor, we tested the results more formally through LRTs. In row 5 of table 2 we reject powerfully the null hypothesis that those with *normalistas* and those with either high school diploma plus one year of college earn the same log-wage. In row 6 we test the analogous hypothesis for the actual wage by imposing the relevant restrictions on g_2 and reject it. The consistency of these results with the predictions of the job specific diploma hypothesis is almost uncanny. *Normalistas*, which are required for teaching jobs whose wages are downwards rigid, receive an abnormally high rate of return. The rate of return to technical high-school degrees, which are tied to jobs whose wages are more competitive, is the same as that to formal high-school. Job-specific diplomas can generate sheepskins if, and only if, the wages paid by these jobs are downwards rigid.

Our fourth set of hypothesis tests seeks sheepskin effects. In the 'logged earnings' column of table 1, we find strong evidence of sheepskin effects in the first moments at a primary level, with almost no return for pre-diploma primary years. The diploma is associated with a 15% wage premium, while each non-diploma years boosts wages by a meager 0.6%. The 'a2' column provides compelling evidence of a positive primary sheepskin effect in 2nd moments. In the employment equation there is also evidence of a negative primary sheepskin effect. No other levels of the formal school system carry a sheepskin effect in first or second moments of logged earnings, or in the employment equation.

These results conjure up a rather vivid picture of equilibrium in a labor market segmented between workers with only primary degrees and those without primary degrees. Anecdotal evidence suggests that some employers in Mexico take the receipt of a primary diploma to signal literacy. This may be why many jobs, especially those in the formal sector, require one explicitly. Thus workers without a diploma may be restricted to a heavily contested, more informal market. In the presence of a wage premium for diplomas, equilibrium across the two markets can only be sustained if the with-diploma market experiences higher unemployment – a la Harris and Todaro (1970). Our finding of a positive primary sheepskin in log earnings and a negative sheepskin in employment is highly evocative of such an equilibrium. The positive sheepskin in the variance of log-earnings suggests, reasonably, that the job opportunities available to primary diploma holders are more varied.

The missing piece of the puzzle is the reason why the wage premium is not bid away. Certainly, if primary graduation required arduous or costly efforts a separating equilibrium could be a plausible answer. We feel that labor market distortions may offer better explanations. We have some, albeit weak, evidence that minimum wages bind more tightly for diploma holders. Despite the negligible earnings increases with primary years, non-diploma holders are twice as likely to earn less than the minimum wage of 5 pesos per hour.[12] This could certainly sustain an equilibrium with unemployment and a modest wage premium in the degreed sector. Another possibility is that the small number of workers without primary degrees are somehow constrained by poverty, the absence of adult education programs and the like, from obtaining them, despite the premium they carry.

The finding that there are almost no diploma effects is especially interesting, maybe even paradoxical, in light of figure 5. In this histogram of schooling completion in our sample there is obvious evidence of clustering at diploma years (years 7, 10, 13 and 18). Education *is* obtained in discrete levels, even if the Mincerian returns do not suggest a reason for this. Clustering may be due to changes across diploma years in the direct costs of education or the non-pecuniary benefits of education, neither of which factor into a Mincerian specification.

[12] Of those with no education beyond primary, 889 out of 3367 non-diploma holders earn below the minimum wage, compared to 445 out of 3029 diploma holders.

Finally, we turn to the 'hours' equation. As shown statistically above, the hours worked by men and women differ from each other and correlate differently with marital status. We therefore focus discussion of the length of the work-week on the results derived from the male and female sub-samples separately (figure 3). The length of the work-week for a bachelor fitting our reference profile declines with formal education, from 54.2 hours for the uneducated to 46.9 for a college graduate. For spinsters it rises from 36.4 hours for the uneducated, peaks at 42.7 for those with a high-school diploma, and then falls to 37.6 hours at college graduation. This coincides well with results from Mehta and Villarreal (2003), which documents, using data from 2000, significantly higher returns to primary, junior-high, and high-school education for women. Many Mexican women work in clerical positions which require junior-high or high-school education and regulated work hours. This would explain why women's wages and hours track education as they do, and why the standard deviation of their hours worked falls with education (figure 3).

The mostly inverse relationship between education and hours worked, coupled with significantly negative estimates of r_3 (see figure 4) may suggest that there are basic necessities which workers with lower earnings potential must labor longer hours to afford. Another likely explanation for the negative correlation between the earnings and hours residuals is simply that hours are in the denominator of earnings, so that those who work more than our model predicts will also tend to earn less than predicted per hour.

VI. Conclusions:

We seek to shed more light on Sheepskin effects and Mincerian returns to education. We develop a 3-equation maximum likelihood specification for an extended Mincerian model, with conditional heteroskedasticity. Using this specification and 2002 data from Mexico we pursue three lines of questioning. First, we ask whether signaling is the definitive explanation for sheepskin effects. We propose an alternative, possibly complimentary explanation: sheepskins may result from job-specific diplomas if, and only if, the wage paid by these jobs are downwards rigid. Second, we test for the implications of these two explanations of sheepskin effects in the second moments of the wage distribution. Third, we ask what happens on the quantity abscissa of the labor market (i.e. unemployment) when sheepskins are seen in wages.

We find extremely strong evidence in support of the job-specific diplomas hypothesis in the results of a natural experiment. Despite the job-specificity of technical high school degrees, they confer the same returns as formal high-schools. Teaching degrees, which lead to a specific job with a downwards rigid wage, confer more than twice the return of either high-school degree. We find positive primary sheepskins in both moments of the wage distribution and a negative one on employment, but no other evidence of diploma effects, despite allowing for diploma and recipient heterogeneity. Our results suggest that where labor markets are segmented according to diplomas, equilibrium is restored through unemployment amongst diploma holders. Our results on hours suggest that much of what is observed regarding Mincerian returns may be driven by behavior in the denominator of implicit wages (hours), rather then just their numerator (salaries).

Our work suggests interesting directions for future research. First, our specification allows us to capture imbalances in the distribution of skills through the returns to heterogeneous diplomas.   Thus, by applying it to data from Mexico through the 1990s, we should be able to trace the formation and resolution of imbalances in Mexican human capital markets in response to NAFTA. Second, some of the results regarding conditional heteroskedasticity in the covariance between earnings and employment invite application of a variant of our model to the empirical modeling of job search strategies. Our sense from our results is that such work might demonstrate how household characteristics influence search strategies with implications for persistent inequality.

References:

Arrow, Kenneth, 1973, "Higher Education as a Filter", *Journal of Political Economy*, 2(3), 193-196.

Azariadis, Costas and A. Drezen, 1990, "Threshold Externalities in Economic Development", *Quarterly Journal of Economics*, 501-26.

Becker, Gary S., 1967, *Human Capital and the Personal Distribution of Income*, Ann Arbor Michigan: University of Michigan Press.

Belman, Dale and John S. Heywood, 1991, "Sheepskin Effects in the returns to education: An Examination of Women and Minorities", *Review of Economics and Statistics*, 73, 720-724.

Card, David, 2001, "Estimating The Return to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, 69(5), 1127-1160.

Chiswick, Barry R., 1973, "Schooling, Screening and Income", in *Does College Matter?*, Lewis Solmon and Paul Taubman (eds.), Academic Press, New York.

Fortin and Lemieux, 1997, "Institutional Changes and Rising Wage Inequality,", *Journal of Economic Perspectives*, 11, 75-96.

Galor, Oded and Joseph Zeira, 1993, "Income Distribution and Macroeconomics", *Review of Economic Studies*, 60: 35-52.

Goldberger, Arthur, 1991, *A Course in Econometrics*, Cambridge, Massachusetts, Harvard University Press.

Harris and Todaro, 1970, "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 40, 126-42.

Heckman, James, 1974, "Shadow wages, market prices, and labor supply," *Econometrica*, 42, 679-94.

Hungerford, Thomas and Gary Solon, 1987, "Sheepskin Effects in the Returns to Education, *Review of Economics and Statistics*, 69, 175-177.

Jaeger, David and Marianne Page, 1996, "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education", *Review of Economics and Statistics*, 78, 733-740.

Ljungqvist, Lars, 1993, "Economic Underdevelopment: The case of a missing market for human capital", *Journal of Development Economics*, 40, 219-39.

Mehta, Aashish and Hector Villarreal, 2003, "Mincerian Returns to Education in Mexico: Sheepskins and Sources of Heterogeneity", University of Wisconsin – Madison, Working Paper.

Mincer, Jacob, 1974, *Schooling, Experience and Earnings*, New York: Columbia University Press.

Mwabu, Germano and T. Paul Schultz, 1996, "Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa", AEA Papers and Proceedings, 82(2), 335-339.

Psacharopoulos, George, 1985, "Returns to Education: A further International Update and Implications" *Journal of Human Resources* 20(4), 583-604.

Riley, John G., 1979, "Testing the Educational Screening Hypothesis", *Journal of Political Economy*, 87(5), S227-S252.

Smith, Paula A. and Michael R. Metzger, 1998, "The Return to Education: Street Vendors in Mexico", *World Development*, 26(2), 289-296.

Spence, A. M., 1974, *Market Signaling*, Harvard University Press, Cambridge .

Appendix A: Derivation of the likelihood function.

The sample is divided between those members of the labor force who are employed ($z_i=1$), and those who are not ($z_i=0$). Hence, if *f( )* denotes the distribution of potential hours and log earnings conditional on employment, the log-likelihood function is of the form:

(A1) $$LLF = \sum_{z_i=0} \ln\left(\Pr\left(z_i = 0\right)\right) + \sum_{z_i=1} \ln\left\{\Pr\left(z_i = 1\right) f\left(y_i^*, h_i^* \middle| z_i = 1\right)\right\}.$$

We suppress *i* for notational purposes for the rest of the derivation. Let F denote the standard normal cumulative distribution function. As usual:

(A2) $$\Pr\left(z = 0\right) = \Pr\left(z^* \leq 0\right) = \Pr\left(b_y x_y + e \leq 0\right) = \Phi\left(-b_y x_y\right).$$

Further, the joint density of y^*, h^* and z in braces in (A1) can be factored differently, and expressed in terms of the latent z^*, rather than z:

(A3) $$\Pr\left(z = 1\right) f\left(y^*, h^* \middle| z = 1\right) = \Pr\left(z = 1 \middle| y^*, h^*\right) g\left(y^*, h^*\right) = \Pr\left(z^* > 0 \middle| y^*, h^*\right) g\left(y^*, h^*\right),$$

where *g( )* is the joint density of y_i^* and h_i^* only.

Following Goldberger (1991), pp.196-97, our normality assumptions (3) imply that $\left(h^*, y^*\right) \sim N\left(m_1, \Sigma_{11}\right)$ and $z^*\big|_{h^*,y^*} \sim N\left(m_2^*, \Sigma_{22}^*\right)$, where:

(A4a) $m_1 = \begin{bmatrix} b_h x_h \\ b_y x_y \end{bmatrix}$, (A4b) $\Sigma_{11} = \begin{bmatrix} q^2 & r_3 qs \\ r_3 qs & s^2 \end{bmatrix}$

(A4c) $$m_2^* = b_z x_z + \frac{r_1 - r_2 r_3}{\left(1 - r_3^2\right)}\left(\frac{h^* - b_h x_h}{q}\right) + \frac{r_2 - r_1 r_3}{\left(1 - r_3^2\right)}\left(\frac{y^* - b_y x_y}{s}\right) \text{ and}$$

(A4d) $$\Sigma_{22}^* = 1 - \left\{A^2 + 2 r_3 AC + C^2\right\} \big/ \left(1 - r_3^2\right)^2; \ A = \left(r_1 - r_2 r_3\right); \ C = \left(r_2 - r_1 r_3\right).$$

Thus,

(A5) $$\Pr\left(z^* > 0 \middle| y^*, h^*\right) = \Phi\left(m_2^* / \Sigma_{22}^*\right) \text{ and}$$

(A6) $g\left(y^*, h^*\right)$ is the bivariate normal pdf characterized by $\left(m_1, \Sigma_{11}\right)$.

Backwards sequential substitution of (A1)-(A6) yield the log likelihood function.

Appendix B: Derivation of moments of the actual wage distribution

Logged earnings (y_i^*) are distributed over the range (-8,8) according to the pdf:

(B1a) $f(y_i^*) = (1/s_i) f\left(\left(y_i^* - b_y x_{yi}\right)/s_i\right)$ and cdf:

(B1b) $F\left(y_i^*\right) = \Pr\left(Y_i^* \le y_i^*\right) = \Phi\left(\left(y_i^* - b_y x_{yi}\right)/s_i\right)$, where f and Φ are the standard normal pdf and cdf respectively.

Because $w_i^* = \exp\left(y_i^*\right)$, earnings fall in the range (0, 8), with cdf:

(B2a) $G\left(w_i^*\right) = \Pr\left(W_i^* \le w_i^*\right) = \Pr\left(Y_i^* \le \ln\left(w_i^*\right)\right) = \Phi\left(\left(\ln w_i^* - b_y x_{yi}\right)/s_i\right)$, and pdf

(B2b) $g(w_i^*) = \partial G(w_i^*)/\partial w_i^* = \left(w_i^* s_i\right)^{-1} f\left(\left(\ln w_i^* - b_y x_{yi}\right)/s_i\right)$.

Then it is straightforward in principle to solve the integrals for:

(B3) $E\left(w_i^*\right) = \int_0^\infty w_i^* g\left(w_i^*\right) dw_i^* = \exp\left(b_y x_{yi} + s_i^2/2\right)$, and

(B4) $V\left(w_i^*\right) = \int_0^\infty g\left(w_i^*\right)\left[w_i^* - \exp\left(b_y x_{yi} + s_i^2/2\right)\right]^2 dw_i^* = \exp\left(2b_y x_{yi}\right)\left[\exp\left(2s_i^2\right) - \exp\left(s_i^2\right)\right]$

Table 1. Parameter estimates for the full sample.*

		Equation							
	(n=16675)	Employment		Hours		Logged Earnings		a2	
-	-	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
	Constant	**-0.2603**	0.0892	**47.8348**	0.7032	**1.5335**	0.0416	**0.6451**	0.0285
	Union			**-3.3819**	0.3020	**0.3585**	0.0144	**-0.1352**	0.0083
	Rural	**0.1251**	0.0421	**-1.3226**	0.2950	**-0.4213**	0.0200	-0.0217	0.0306
	Male	**-0.3234**	0.0426	**2.0179**	0.3601	**0.1833**	0.0141	**0.0706**	0.0187
Intercept	North	**-0.1516**	0.0368	-0.4552	0.2377	**0.1187**	0.0129		
Shifters	South	**0.1020**	0.0438	**2.1727**	0.2445	**-0.2317**	0.0141		
	Married	**0.8161**	0.0876	**-4.5950**	0.3589				
	Married Male	0.1236	0.0949	**8.5991**	0.4402				
	Experience	**0.0216**	0.0047	**0.1110**	0.0306	**0.0384**	0.0018	**-0.0118**	0.0009
	Experience squared	-0.0001	0.0001	**-0.0032**	0.0005	**-0.0005**	0.0000	**0.0002**	0.0000
	Primary years	**0.3016**	0.0138	**-0.3746**	0.0807	0.0063	0.0073	**-0.0230**	0.0047
	Primary diploma	**-0.6232**	0.0854			**0.1428**	0.0308	**0.0802**	0.0196
	Junior-high years	0.0096	0.0509	0.1155	0.1011	**0.0657**	0.0156	0.0031	0.0081
	Junior-high diploma	0.1169	0.1423			-0.0150	0.0436	-0.0190	0.0224
Schooling	Formal high-school years	0.0220	0.0850	-0.1996	0.1388	**0.1167**	0.0190	**0.0316**	0.0108
Effects	Formal high-School diploma	-0.4006	0.2506			-0.0261	0.0582	0.0302	0.0328
	3 year technical high-school degree	**-0.1556**	0.0753	-0.3506	0.4156	**0.3286**	0.0254	**0.1620**	0.0141
	4 year normalista					**0.8746**	0.0558	**0.0876**	0.0316
	College years, post formal			**-0.9006**	0.1027	**0.1589**	0.0104	0.0099	0.0067
	College diploma, post formal					-0.0019	0.0457	-0.0555	0.0308
	College years, post technical			-0.1257	0.1989	**0.1546**	0.0242	-0.0419	0.0218
	College years, post formal			**-8.7582**	1.0458	-0.1200	0.1290	0.1301	0.1144
	College years, all HS graduates	0.0320	0.0279						
	College diploma, all HS graduates	-0.0072	0.1338						
		a1		a3		a4		a5	
-	-	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
	Constant	**13.2330**	0.3220	**-10.6251**	0.6839	**-0.4810**	0.0342	**-0.5133**	0.0227
	Union	**-1.2439**	0.1575					-0.0220	0.0113
	Rural	**1.2803**	0.1686	0.1541	0.6405	**-0.2984**	0.0311	-0.0399	0.0268
	Male	**-0.7619**	0.1837	**5.2158**	0.5177	**0.0980**	0.0242	**0.1600**	0.0171
	Normalista	0.8919	0.6274					**-0.1806**	0.0406
	Schooling	**-0.0920**	0.0225	**0.7215**	0.0473	0.0040	0.0028	**0.0222**	0.0018

* **Bold** coefficients indicate 95% significance.

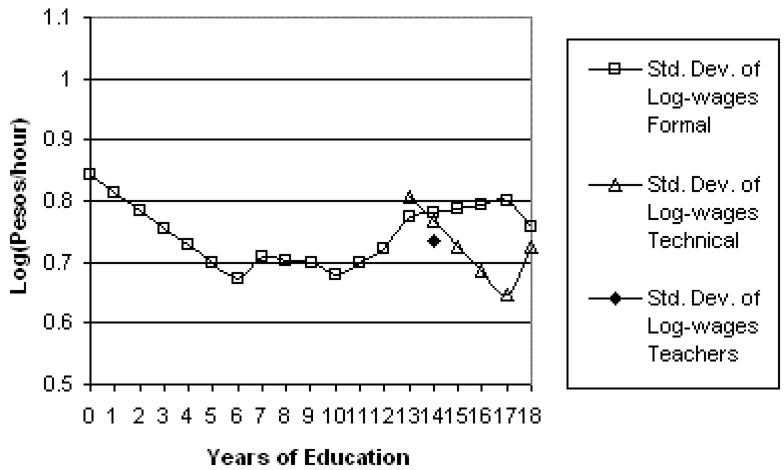

Figure 1: Std. Dev. of Log-wages

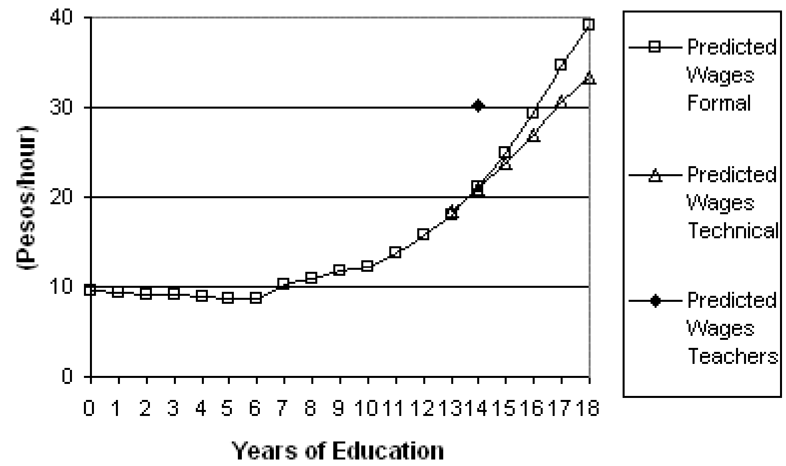

Figure 2: Predicted Wages

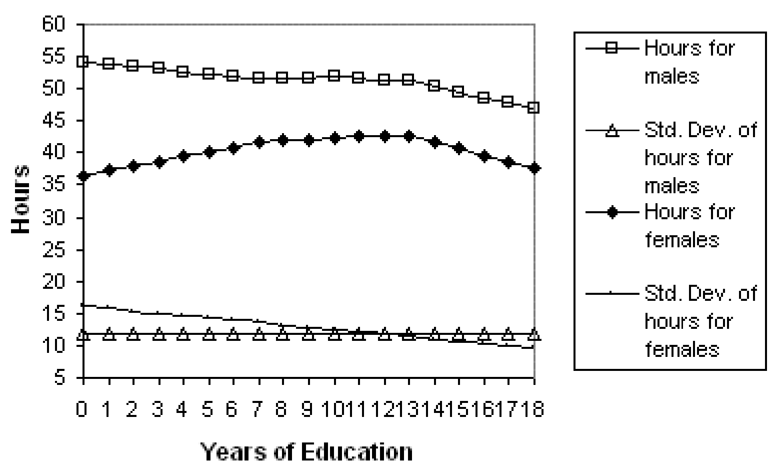

Figure 3: Predicted Hours and their standard deviations

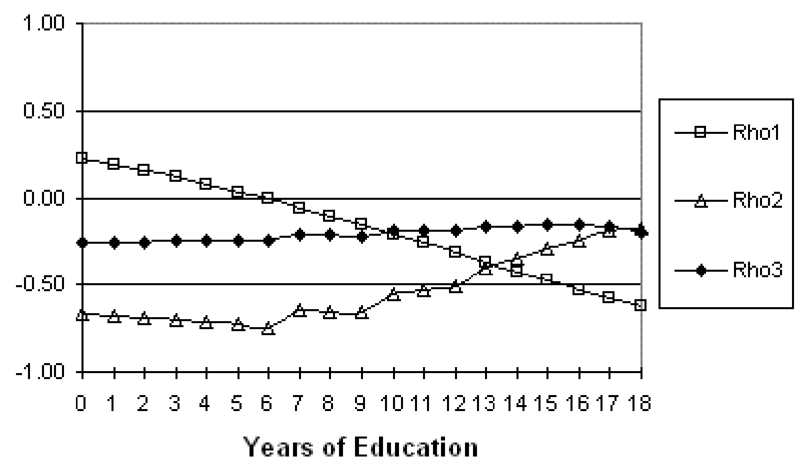

Figure 4: Correlation between residuals

*Table 2. Likelihood Ratio Test Results: Comparisons to the unrestricted model:***

	Null Hypothesis	Log Likelihood Value	LRT statistic	Degrees of Freedom	p-value
1.	Homoskedasticity	-80,163.46	1343.10	34	0.0000
2.	No self selection	-79,580.99	178.15	4	0.0000
3.	Formal - Technical log wage equality	-79,491.92	0.02	1	0.8784
4.	Formal - Technical wage equality	-79,493.10	2.38	2	0.3042
5.	Formal - Technical - Normalista log wage equality	-79,516.03	48.24	2	0.0000
6.	Formal - Technical - Normalista log wage equality	-79,516.20	48.58	4	0.0000

**Log likelihood value = -79,491.9100

Key

Degree Year	Years of Schooling
No education	0
Kindergarten	1
Primary, year 1	2
Primary, year 2	3
Primary, year 3	4
Primary, year 4	5
Primary, year 5	6
Primary Graduation	7
Junior-high, year 1	8
Junior-high, year 2	9
Junior-high Graduation	10
Formal High-school, year 1	11
Formal High-school, year 2	12
Formal High-school Graduation	13
Technical High-School Degree	13
Normalista	14
College, year 1	14
College, year 2	15
College, year 3	16
College, year 4	17
College Graduation	18

Figure 5: Distribution of completed years of schooling.

(Normalista recipients and those with only technical high-school degrees excluded.)